A/ Box Sequence

PATENT

Docket No. GM10127

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of: Nicola G. Wallis, et al

For: Histidine Kinase

Assistant Commissioner for Patents
Box Patent Application
Washington, D.C. 20231

COVER SHEET FOR FILING PATENT APPLICATION

Enclosed herewith for filing in the United States Patent and Trademark Office are the Specification and Declaration and Power of Attorney for the patent application identified below:

1. The names of the inventors are: Nicola G. Wallis, Lisa K. Shilling, Jeffrey L. Mooney, Christine Debouck, YiYi Zhong, Deborah D. Jaworski, Min Wang, and John P. Throup

2. The addresses of the inventors are:

Nicola Gail Wallis
K203, 219 Sugartown Rd,
Wayne, PA19087; USA
(Citizen of the United Kingdom)

Lisa Kathleen Shilling
6 East Park Rd
Newtown, Pa 18940; USA
(Citizen of the United States of America)

Jeffrey L Mooney
143 Winged Foot court
Limerick, PA 19468; USA
(Citizen of the United States of America)

Christine Debouck
667 Pugh Rd
Wayne, PA 19087; USA
(Citizen of Belgium)

Yi Yi Zhong
2828 Egypt Road, Apt. D305
Audubon, PA 19403; USA
(Citizen of China)

Deborah Dee Jaworski
214 Willowbrook, Dr.
Norristown, PA 19403; USA
(Citizen of the United States of America)

Min Wang
302 Centennial Drive
Blue Bell, PA 19422; USA
(Citizen of China)

John Peter Throup
506 Brooke Drive
Royersford, PA 19468; USA
(Citizen of the United Kingdom)

3. The title of the invention is: Histidine Kinase

4. The name, registration and telephone number of the attorney is:
Attorney: Stephen T. Falk
Registration No. 36,795
Telephone No. 215-994-2488

5. The Docket Number used to identify this application is: GM10127

6. The correspondence address for this application is:
Stephen T. Falk
Dechert, Price & Rhoads
4000 Bell Atlantic Tower, 1717 Arch Street
Philadelphia, PA 19103-2793

This invention was not made by an agency of the United States Government, or under contract with an agency of the United States Government.

There are 47 pages of specification, 0 sheets of drawings, 24 claims and a diskette containing a CRF copy of the Sequence Listing accompanying this cover sheet.

Stephen T. Falk
Registration No. 36,795
Dechert, Price & Rhoads
4000 Bell Atlantic Tower
1717 Arch Street
Philadelphia, PA 19103-2793, USA
Phone: 215/994-2488

Date:January 13, 1998

CERTIFICATION UNDER 37 CFR 1.10

I hereby certify that this correspondence and the documents referred to as attached therein are being deposited with the United States Postal Service on **January 13, 1998**, in an envelope as "EXPRESS MAIL POST OFFICE TO ADDRESSEE" service under 37 CFR 1.10, Mailing Label Number **EE 065 729 589 US** addressed to: Assistant Commissioner for Patents, Washington, D.C. 20231.

Robert Ballarini

Docket No. GM10127

INVENTORS:

Nicola Gail Wallis
K203, 219 Sugartown Rd,
Wayne, PA19087; USA
(Citizen of the United Kingdom)

Lisa Kathleen Shilling
6 East Park Rd
Newtown, Pa 18940; USA
(Citizen of the United States of America)

Jeffrey L Mooney
143 Winged Foot court
Limerick, PA 19468; USA
(Citizen of the United States of America)

Christine Debouck
667 Pugh Rd
Wayne, PA 19087; USA
(Citizen of Belgium)

Yi Yi Zhong
2828 Egypt Road, Apt. D305
Audubon, PA 19403; USA
(Citizen of China)

Deborah Dee Jaworski
214 Willowbrook, Dr.
Norristown, PA 19403; USA
(Citizen of the United States of America)

Min Wang
302 Centennial Drive
Blue Bell, PA 19422; USA
(Citizen of China)

John Peter Throup
506 Brooke Drive
Royersford, PA 19468; USA
(Citizen of the United Kingdom)

TITLE: Histidine Kinase

Histidine Kinase

FIELD OF THE INVENTION

This invention relates to newly identified polynucleotides and polypeptides, and their production and uses, as well as their variants, agonists and antagonists, and their uses. In particular, the invention relates to novel polynucleotides and polypeptides of the Histidine Kinase family, hereinafter referred to as "Histidine Kinase".

BACKGROUND OF THE INVENTION

It is particularly preferred to employ Staphylococcal genes and gene products as targets for the development of antibiotics. The Staphylococci make up a medically important genera of microbes. They are known to produce two types of disease, invasive and toxigenic. Invasive infections are characterized generally by abscess formation effecting both skin surfaces and deep tissues. *Staphylococcus aureus* is the second leading cause of bacteremia in cancer patients. Osteomyelitis, septic arthritis, septic thrombophlebitis and acute bacterial endocarditis are also relatively common. There are at least three clinical conditions resulting from the toxigenic properties of Staphylococci. The manifestation of these diseases result from the actions of exotoxins as opposed to tissue invasion and bacteremia. These conditions include: Staphylococcal food poisoning, scalded skin syndrome and toxic shock syndrome.

The frequency of *Staphylococcus aureus* infections has risen dramatically in the past few decades. This has been attributed to the emergence of multiply antibiotic resistant strains and an increasing population of people with weakened immune systems. It is no longer uncommon to isolate *Staphylococcus aureus* strains which are resistant to some or all of the standard antibiotics. This phenomenon has created a demand for both new anti-microbial agents, vaccines, and diagnostic tests for this organism.

Many two component signal transduction systems (TCSTS) have been identified in bacteria (Stock, J. B., Ninfa, A.J. & Stock, A.M.(1989) Microbiol. Rev. 53, 450-490). These are involved in the bacterium's ability to monitor its surroundings and adapt to changes in its

environment. Several of these bacterial TCSTS are involved in virulence and bacterial pathogenesis within the host.

Histidine kinases are components of the TCSTS which autophosphorylate at a histidine residue. The phosphate group is then transferred to the cognate response regulator. The histidine kinases have five short conserved amino acid sequences (Stock, J. B., Ninfa, A.J.& Stock, A.M.(1989) Microbiol. Rev. 53, 450-490, Swanson, R.V., Alex, L.A. & Simon, M.I.(1994) TIBS 19 485-491). These are the histidine residue, which is phosphorylated, followed after approximately 100 residues by a conserved asparagine residue. After another 15 to 45 residues a DXGXG motif is found, followed by a FXXF motif after another 10-20 residues. 10-20 residues further on another glycine motif, GXG is found. The two glycine motifs are thought to be involved in nucleotide binding. This family of histidine kinases includes KinA protein from *Lactococcus lactis cremoris* (O'Connell-Motherway, M., Fitzgerald, G.F. & van Sinderen, D. (1997) Appl. Environ. Microbiol. 63 2454-2459).

Response regulators are components of the TCSTS. These proteins are phosphorylated from histidine kinases and in turn once phosphorylated affect the response, often through a DNA binding domain becoming activated. The response regulators are characterized by a conserved N-terminal domain of approximately 100 amino acids. The N-terminal domains of response regulators as well as retaining five functionally important residues, corresponding to the residues D12, D13, D57, T87, K109 in CheY (Matsumura, P., Rydel, J.J., Linzmeier, R. & Vacante, D. (1984) J. Bacteriol. 160, 36-41), have conserved structural features (Volz, K. (1993) Biochemistry 32, 11741-11753). The 3-dimensional structures of CheY from *Salmonella typhimurium* (Stock, A.M., Mottonen, J.M., Stock, J.B.& Schutt, ,C.E. (1989) *Nature*, **337**, 745-749) and *Escherichia coli* (Volz, K. & Matsumura, P. (1991) *J. Biol. Chem.* **266,** 15511-15519) and the N-terminal domain of nitrogen regulatory protein C from *S.typhimurium* (Volkman, B.F., Nohaile, M.J., Amy, N.K., Kustu, S. & Wemmer, D.E. (1995) Biochemistry, 34 1413-1424), are available, as well as the secondary structure of SpoOF from *Bacillus subtilis* (Feher, V.A., Zapf, J.W., Hoch, J.A., Dahlquist, F.W., Whiteley, J.M. & Cavanagh, J. (1995) Protein Science, 4, 1801-1814). These structures have a $(a/b)_5$ fold. Several structural residues are conserved between different response regulator sequences, specifically hydrophobic residues within the β-sheet hydrophobic core and sites from the a-helices.

Among the processes regulated by TCSTS are production of virulence factors, motility, antibiotic resistance and cell replication. Inhibitors of TCSTS proteins would prevent the bacterium from establishing and maintaining infection of the host by preventing it from producing the necessary factors for pathogenesis and thereby have utility in anti-bacterial therapy.

Clearly, there exists a need for factors, such as the Histidine Kinase embodiments of the invention, that have a present benefit of being useful to screen compounds for antibiotic activity. Such factors are also useful to determine their role in pathogenesis of infection, dysfunction and disease. There is also a need for identification and characterization of such factors and their antagonists and agonists to find ways to prevent, ameliorate or correct such infection, dysfunction and disease.

Certain of the polypeptides of the invention possess amino acid sequence homology to a known KinA from *Lactococcus lactis cremoris* protein. (O'Connell-Motherway, M., Fitzgerald, G.F. & van Sinderen, D. (1997) Appl. Environ. Microbiol. 63 2454-2459; U81166 from Genbank.)

SUMMARY OF THE INVENTION

It is an object of the invention to provide polypeptides that have been identified as novel Histidine Kinase polypeptides by homology between the amino acid sequence set out in Table 1 [SEQ ID NO: 2] and a known amino acid sequence or sequences of other proteins such as KinA from *Lactococcus lactis cremoris* protein.

It is a further object of the invention to provide polynucleotides that encode Histidine Kinase polypeptides, particularly polynucleotides that encode the polypeptide herein designated Histidine Kinase.

In a particularly preferred embodiment of the invention, the polynucleotide comprises a region encoding Histidine Kinase polypeptides comprising a sequence set out in Table 1 [SEQ ID NO:1] which includes a full length gene, or a variant thereof.

In another particularly preferred embodiment of the invention, there is a novel Histidine Kinase protein from *Staphylococcus aureus* comprising the amino acid sequence of Table 1 [SEQ ID NO:2], or a variant thereof.

As a further aspect of the invention, there are provided isolated nucleic acid molecules encoding Histidine Kinase, particularly *Staphylococcus aureus* Histidine Kinase, including mRNAs, cDNAs, genomic DNAs. Further embodiments of the invention include biologically, diagnostically, prophylactically, clinically or therapeutically useful variants thereof, and compositions comprising the same.

In accordance with another aspect of the invention, there is provided the use of a polynucleotide of the invention for therapeutic or prophylactic purposes, in particular genetic immunization. Among the particularly preferred embodiments of the invention are naturally occurring allelic variants of Histidine Kinase and polypeptides encoded thereby.

In another aspect of the invention, there are provided novel polypeptides of *Staphylococcus aureus* referred to herein as Histidine Kinase as well as biologically, diagnostically, prophylactically, clinically or therapeutically useful variants thereof, and compositions comprising the same.

Among the particularly preferred embodiments of the invention are variants of Histidine Kinase polypeptide encoded by naturally occurring alleles of the Histidine Kinase gene.

In a preferred embodiment of the invention, there are provided methods for producing the aforementioned Histidine Kinase polypeptides.

In accordance with yet another aspect of the invention, there are provided inhibitors to such polypeptides, useful as antibacterial agents, including, for example, antibodies.

In accordance with certain preferred embodiments of the invention, there are provided products, compositions and methods for assessing Histidine Kinase expression, treating disease, assaying genetic variation, and administering a Histidine Kinase polypeptide or polynucleotide to an organism to raise an immunological response against a bacteria, especially a *Staphylococcus aureus* bacteria.

In accordance with certain preferred embodiments of this and other aspects of the invention, there are provided polynucleotides that hybridize to Histidine Kinase polynucleotide sequences, particularly under stringent conditions.

In certain preferred embodiments of the invention, there are provided antibodies against Histidine Kinase polypeptides.

In other embodiments of the invention, there are provided methods for identifying compounds which bind to or otherwise interact with and inhibit or activate an activity of a

polypeptide or polynucleotide of the invention comprising: contacting a polypeptide or polynucleotide of the invention with a compound to be screened under conditions to permit binding to or other interaction between the compound and the polypeptide or polynucleotide to assess the binding to or other interaction with the compound, such binding or interaction being associated with a second component capable of providing a detectable signal in response to the binding or interaction of the polypeptide or polynucleotide with the compound; and determining whether the compound binds to or otherwise interacts with and activates or inhibits an activity of the polypeptide or polynucleotide by detecting the presence or absence of a signal generated from the binding or interaction of the compound with the polypeptide or polynucleotide.

In accordance with yet another aspect of the invention, there are provided Histidine Kinase agonists and antagonists, preferably bacteriostatic or bacteriocidal agonists and antagonists.

In a further aspect of the invention, there are provided compositions comprising a Histidine Kinase polynucleotide or a Histidine Kinase polypeptide for administration to a cell or to a multicellular organism.

Various changes and modifications within the spirit and scope of the disclosed invention will become readily apparent to those skilled in the art from reading the following descriptions and from reading the other parts of the present disclosure.

DESCRIPTION OF THE INVENTION

The invention relates to novel Histidine Kinase polypeptides and polynucleotides as described in greater detail below. In particular, the invention relates to polypeptides and polynucleotides of a novel Histidine Kinase of *Staphylococcus aureus*, which is related by amino acid sequence homology to KinA from *Lactococcus lactis cremoris* polypeptide. The invention relates especially to Histidine Kinase having the nucleotide and amino acid sequences set out in Table 1 as SEQ ID NO: 1 and SEQ ID NO: 2 respectively .

TABLE 1

Histidine Kinase Polynucleotide and Polypeptide Sequences

(A) Sequences from *Staphylococcus aureus* Histidine Kinase polynucleotide sequence [SEQ ID NO:1].

5'-TAATTTAAAAAGCAACTATTGTATAGAAAAATACAAAATTTAAAATATATTACCTTATTA
GAAAAAGTCGTAATATGAGGTGTACAAATGACGCAAATTTTAATAGTAGAAGATGAACAA
AACTTAGCAAGATTTCTTGAATTGGAACTCACACATGAAAATTACAATGTGGACACAGAG
TATGATGGACAAGACGGTTTAGATAAAGCGCTTAGCCATTACTATGATTTAATCATATTA
GATTTAATGTTGCCGTCAATTAATGGCTTAGAAATTTGTCGCAAAATTAGACAACAACAA
TCTACACCTATCATTATAATTACAGCGAAAAGTGATACGTATGACAAAGTTGCTGGGCTT
GATTACGGTGCAGACGATTATATAGTTAAGCCGTTTGATATTGAAGAACTTTTAGCAAGA
ATTCGTGCAATTTTACGTCGTCAGCCACAAAAGGATATTATCGATGTCAACGGTATTACA
ATTGATAAGAACGCTTTTAAAGTGACGGTAAATGGCGCAGAAATTGAATTAACAAAAACA
GAGTATGATTTACTATATCTTCTAGCTGAAAATAAAAACCATGTTATGCAACGGGAACAA
ATTTTAAATCATGTATGGGGTTATAATAGTGAAGTAGAAACAAATGTCGTAGATGTTTAT
ATAAGATATTTACGAAACAAGTTAAAACCATACGATCGTGACAAAATGATTGAAACAGTT
CGTGGCGTTGGGTATGTGATACGATGACAAAACGTAAATTGCGCAATAACTGGATTATTG
TTACCACGATGATTACGTTTGTCACGATATTTTTGTTTTGTTTAATTATTATTTTTTTCT
TGAAAGATACACTGCATAATAGTGAGCTTGATGATGCAGAACGAAGCTCAAGCGATATTA
ATAATTTATTTCATTCTAAGCCTGTTAAAGATATATCTGCATTAGACTTGAATGCATCTT
TAGGTAATTTTCAAGAGATAATTATTTATGATGAGCATAATAATAAATTATTTGAGACAT
CGAATGATAACACAGTGAGAGTTGAACCAGGTTATGAACACCGTTATTTTGACCGCGTAA
TAAAAAAACGCTATAAAGGCATTGAATATTTAATTATTAAAGAACCAATTACAACGCAAG
ATTTCAAAGGGTATAGCTTGTTAATTCATTCACTAGAAAATTATGATAACATCGTAAAAT
CATTGTATATCATTGCGCTGGCATTTGGAGTGATTGCAACAATTATAACTGCCACAATCA
GTTATGTATTTTCAACACAAATTACTAAACCGCTTGTCAGTTTATCAAATAAAATGATTG
AGATTCGACGAGATGGTTTTCAAAATAAATTGCAATTAAATACAAATTATGAAGAAATAG
ATAATTTAGCAAATACGTTTAATGAGATGATGAGCCAAATTGAAGAATCATTTAATCAAC
AAAGACAATTTGTTGAAGATGCGTCACATGAATTACGAACACCATTACAAATTATTCAAG
GTCATTTAAATTTGATTCAGCGATGGGGAAAAAAAGACCCAGCAGTATTAGAAGAATCGT
TAAATATTTCTATTGAAGAAATGAATCGTATCATAAAATTAGTCGAAGAATTACTTGAAT
TGACTAAAGGAGATGTAAATGACATTTCTTCTGAAGCGCAGACCGTGCATATTAATGATG
AAATTCGCTCGCGAATACACTCATTAAAACAATTGCATCCTGATTATCAATTTGATACGG
ATCTGACATCTAAAAATCTAGAAATTAAAATGAAACCTCATCAATTCGAACAATTATTTT
TAATCTTTATTGATAATGCAATCAAATATGATGTGAAGAATAAGAAAATTAAAGTTAAGA
CAAGGTTAAAAAATAAGCAAAAAATAATTGAAATTACAGATCATGGAATTGGTATTCCAG
AGGAAGATCAAGATTTCATTTTTGATCGCTTTTATCGAGTGGATAAATCTCGTTCAAGAA
GTCAAGGCGGTAATGGACTCGGATTATCTATTGCTCAAAAAATCATTCAATTAAACGGAG
GATCGATTAAAATTAAAAGTGAAATTAATAAAGGAACAACGTTTAAAATCATATTTTAAT
CATGTCTGAGACGTCAATCAAAGTCATAGGATCAATTTTTTAAGTACACATTAGCTGTGA

CTAATGTATAAGAACAACTATAAAACAAATAAACAGTGGTT-3'

(B) *Staphylococcus aureus* Histidine Kinase polypeptide sequence deduced from the polynucleotide sequence in this table [SEQ ID NO:2].

NH_2-MTKRKLRNNWIIVTTMITFVTIFLFCLIIIFFLKDTLHNSELDDAERSSSDINNLFHSKP
VKDISALDLNASLGNFQEIIIYDEHNNKLFETSNDNTVRVEPGYEHRYFDRVIKKRYKGI
EYLIIKEPITTQDFKGYSLLIHSLENYDNIVKSLYIIALAFGVIATIITATISYVFSTQI
TKPLVSLSNKMIEIRRDGFQNKLQLNTNYEEIDNLANTFNEMMSQIEESFNQQRQFVEDA
SHELRTPLQIIQGHLNLIQRWGKKDPAVLEESLNISIEEMNRIIKLVEELLELTKGDVND
ISSEAQTVHINDEIRSRIHSLKQLHPDYQFDTDLTSKNLEIKMKPHQFEQLFLIFIDNAI
KYDVKNKKIKVKTRLKNKQKIIEITDHGIGIPEEDQDFIFDRFYRVDKSRSRSQGGNGLG
LSIAQKIIQLNGGSIKIKSEINKGTTFKIIF-COOH

(C) Polynucleotide sequences comprising *S.aureus* response regulator cognate to the histidine kinase of the invention [SEQ ID NO 3]

ATTTACGTTT TGTCATCGTA TCACATACCC AACGCCACGA ACTGTTTCAA TCATTTTGTC 60

ACGATCGTAT GGTTTTAACT TGTTTCGTAA ATATCTTATA TAAACATCTA CGACATTTGT 120

TTCTACTTCA CTATTATAAC CCCATACATG ATTTAAAATT TGTTCCCGTT GCATAACATG 180

GTTTTTATTT TCAGCTAGAA GATATAGTAA ATCATACTCT GTTTTTGTTA ATTCAATTTC 240

TGCGCCATTT ACCGTCACTT TAAAAGCGTT CTTATCAATT GTAATACCGT TGACATCGAT 300

AATATCCTTT TGTGGCTGAC GACGTAAAAT TGCACGAATT CTTGCTAAAA GTTCTTCAAT 360

ATCAAACGGC TTAACTATAT AATCGTCTGC ACCGTAATCA AGCCCAGCAA CTTTGTCATA 420

CGTATCACTT TTCGCTGTAA TTATAATGAT AGGTGTAGAT TGTTGTTGTC TAATTTTGCG 480

ACAAATTTCT AAGCCATTAA TTGACGGCAA CATTAAATCT AATATGATTA AATCATAGTA 540

ATGGCTAAGC GCTTTATCTA AACCGTCTTG TCCATCATAC TCTGTGTCCA CATTGTAATT 600

TTCATGTGTG AGTTCCAATT CAAGAAATCT TGCTAAGTTT TGTTCATCTT CTACTATTAA 660

AATTTGCGTC ATTTGTACAC CTCATATTAC GACTTTTTCT AATAAGGTAA TATATTTTAA 720

ATTTTGTATT TTTCTA 736

(D) Polypeptide sequence of *Staphylococcus aureus* response regulator cognate of the histidine kinase of the invention [SEQID NO 4]

1 MTQILIVEDE QNLARFLELE LTHENYNVDT EYDGQDGLDK ALSHYYDLII
51 LDLMLPSING LEICRKIRQQ QSTPIIIITA KSDTYDKVAG LDYGADDYIV
101 KPFDIEELLA RIRAILRRQP QKDIIDVNGI TIDKNAFKVT VNGAEIELTK
151 TEYDLLYLLA ENKNHVMQRE QILNHVWGYN SEVETNVVDV YIRYLRNKLK
201 PYDRDKMIET VRGVGYVIR

Deposited materials

A deposit containing a *Staphylococcus aureus* WCUH 29 strain has been deposited with the National Collections of Industrial and Marine Bacteria Ltd. (herein "NCIMB"), 23 St. Machar Drive, Aberdeen AB2 1RY, Scotland on 11 September 1995 and assigned NCIMB Deposit No. 40771, and referred to as *Staphylococcus aureus* WCUH29 on deposit. The *Staphylococcus aureus* strain deposit is referred to herein as "the deposited strain" or as "the DNA of the deposited strain."

The deposited strain contains the full length Histidine Kinase gene. The sequence of the polynucleotides contained in the deposited strain, as well as the amino acid sequence of the polypeptide encoded thereby, are controlling in the event of any conflict with any description of sequences herein.

The deposit of the deposited strain has been made under the terms of the Budapest Treaty on the International Recognition of the Deposit of Micro-organisms for Purposes of Patent Procedure. The strain will be irrevocably and without restriction or condition released to the public upon the issuance of a patent. The deposited strain is provided merely as convenience to those of skill in the art and is not an admission that a deposit is required for enablement, such as that required under 35 U.S.C. §112.

A license may be required to make, use or sell the deposited strain, and compounds derived therefrom, and no such license is hereby granted.

One aspect of the invention there is provided an isolated nucleic acid molecule encoding a mature polypeptide expressible by the *Staphylococcus aureus* WCUH 29 strain contained in the deposited strain. Further provided by the invention are Histidine Kinase nucleotide sequences of the DNA in the deposited strain and amino acid sequences encoded thereby. Also provided by the invention are Histidine Kinase polypeptide sequences isolated from the deposited strain and amino acid sequences derived therefrom.

Polypeptides

The polypeptides of the invention include a polypeptide of Table 1 [SEQ ID NO:2] (in particular the mature polypeptide) as well as polypeptides and fragments, particularly those which have the biological activity of Histidine Kinase, and also those which have at least 70% identity to a polypeptide of Table 1 [SEQ ID NO:1]or the relevant portion, preferably at least 80% identity to

a polypeptide of Table 1 [SEQ ID NO:2and more preferably at least 90% similarity (more preferably at least 90% identity) to a polypeptide of Table 1 [SEQ ID NO:2] and still more preferably at least 95% similarity (still more preferably at least 95% identity) to a polypeptide of Table 1 [SEQ ID NO:2] and also include portions of such polypeptides with such portion of the polypeptide generally containing at least 30 amino acids and more preferably at least 50 amino acids.

The invention also includes polypeptides of the formula:

$$X\text{-}(R_1)_m\text{-}(R_2)\text{-}(R_3)_n\text{-}Y$$

wherein, at the amino terminus, X is hydrogen, and at the carboxyl terminus, Y is hydrogen or a metal, R_1 and R_3 are any amino acid residue, m is an integer between 1 and 1000 or zero, n is an integer between 1 and 1000 or zero, and R_2 is an amino acid sequence of the invention, particularly an amino acid sequence selected from Table 1. In the formula above R_2 is oriented so that its amino terminal residue is at the left, bound to R_1, and its carboxy terminal residue is at the right, bound to R_3. Any stretch of amino acid residues denoted by either R group, where m and/or n is greater than 1, may be either a heteropolymer or a homopolymer, preferably a heteropolymer.

A fragment is a variant polypeptide having an amino acid sequence that entirely is the same as part but not all of the amino acid sequence of the aforementioned polypeptides. As with Histidine Kinase polypeptides fragments may be "free-standing," or comprised within a larger polypeptide of which they form a part or region, most preferably as a single continuous region, a single larger polypeptide.

Preferred fragments include, for example, truncation polypeptides having a portion of an amino acid sequence of Table 1 [SEQ ID NO:2], or of variants thereof, such as a continuous series of residues that includes the amino terminus, or a continuous series of residues that includes the carboxyl terminus. Degradation forms of the polypeptides of the invention in a host cell, particularly a *Staphylococcus aureus*, are also preferred. Further preferred are fragments characterized by structural or functional attributes such as fragments that comprise alpha-helix and alpha-helix forming regions, beta-sheet and beta-sheet-forming regions, turn and turn-forming regions, coil and coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions, substrate binding region, and high antigenic index regions.

Also preferred are biologically active fragments which are those fragments that mediate activities of Histidine Kinase, including those with a similar activity or an improved activity, or with a decreased undesirable activity. Also included are those fragments that are antigenic or immunogenic in an animal, especially in a human. Particularly preferred are fragments comprising receptors or domains of enzymes that confer a function essential for viability of *Staphylococcus aureus* or the ability to initiate, or maintain cause disease in an individual, particularly a human.

Variants that are fragments of the polypeptides of the invention may be employed for producing the corresponding full-length polypeptide by peptide synthesis; therefore, these variants may be employed as intermediates for producing the full-length polypeptides of the invention.

In addition to the standard single and triple letter representations for amino acids, the term "X" or "Xaa" may also be used in describing certain polypeptides of the invention. "X" and "Xaa" mean that any of the twenty naturally occuring amino acids may appear at such a designated position in the polypeptide sequence.

Polynucleotides

Another aspect of the invention relates to isolated polynucleotides, including the full length gene, that encode the Histidine Kinase polypeptide having a deduced amino acid sequence of Table 1 [SEQ ID NO:2] and polynucleotides closely related thereto and variants thereof.

Using the information provided herein, such as a polynucleotide sequence set out in Table 1 [SEQ ID NO:1], a polynucleotide of the invention encoding Histidine Kinase polypeptide may be obtained using standard cloning and screening methods, such as those for cloning and sequencing chromosomal DNA fragments from bacteria using *Staphylococcus aureus* WCUH 29 cells as starting material, followed by obtaining a full length clone. For example, to obtain a polynucleotide sequence of the invention, such as a sequence given in Table 1 [SEQ ID NO:1], typically a library of clones of chromosomal DNA of *Staphylococcus aureus* WCUH 29 in *E.coli* or some other suitable host is probed with a radiolabeled oligonucleotide, preferably a 17-mer or longer, derived from a partial sequence. Clones carrying DNA identical to that of the probe can then be distinguished using stringent conditions. By sequencing the individual clones thus identified with sequencing primers designed from the original sequence it is then possible to extend the sequence in both directions to determine the full gene sequence.

Conveniently, such sequencing is performed using denatured double stranded DNA prepared from a plasmid clone. Suitable techniques are described by Maniatis, T., Fritsch, E.F. and Sambrook et al., *MOLECULAR CLONING, A LABORATORY MANUAL*, 2nd Ed.; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989). (see in particular Screening By Hybridization 1.90 and Sequencing Denatured Double-Stranded DNA Templates 13.70). Illustrative of the invention, the polynucleotide set out in Table 1 [SEQ ID NO:1] was discovered in a DNA library derived from *Staphylococcus aureus* WCUH 29.

The DNA sequence set out in Table 1 [SEQ ID NO:1] contains an open reading frame encoding a protein having about the number of amino acid residues set forth in Table 1 [SEQ ID NO:2] with a deduced molecular weight that can be calculated using amino acid residue molecular weight values well known in the art. The polynucleotide of SEQ ID NO: 1, between nucleotide number 744 and the stop codon which begins at nucleotide number 2097 of SEQ ID NO:1, encodes the polypeptide of SEQ ID NO:2.

Histidine Kinase of the invention is structurally related to other proteins of the Histidine Kinase family.

The invention provides a polynucleotide sequence identical over its entire length to a coding sequence in Table 1 [SEQ ID NO:1]. Also provided by the invention is the coding sequence for the mature polypeptide or a fragment thereof, by itself as well as the coding sequence for the mature polypeptide or a fragment in reading frame with other coding sequence, such as those encoding a leader or secretory sequence, a pre-, or pro- or prepro- protein sequence. The polynucleotide may also contain non-coding sequences, including for example, but not limited to non-coding 5' and 3' sequences, such as the transcribed, non-translated sequences, termination signals, ribosome binding sites, sequences that stabilize mRNA, introns, polyadenylation signals, and additional coding sequence which encode additional amino acids. For example, a marker sequence that facilitates purification of the fused polypeptide can be encoded. In certain embodiments of the invention, the marker sequence is a hexa-histidine peptide, as provided in the pQE vector (Qiagen, Inc.) and described in Gentz *et al.*, *Proc. Natl. Acad. Sci., USA 86:* 821-824 (1989), or an HA tag (Wilson *et al.*, *Cell 37:* 767 (1984). Polynucleotides of the invention also include, but are not limited to, polynucleotides comprising a structural gene and its naturally associated sequences that control gene expression.

A preferred embodiment of the invention is a polynucleotide of comprising nucleotide 744 to the nucleotide immediately upstream of or including nucleotide 2097 set forth in SEQ ID NO:1 of Table 1, both of which encode the Histidine Kinase polypeptide.

The invention also includes polynucleotides of the formula:

$$X\text{-}(R_1)_m\text{-}(R_2)\text{-}(R_3)_n\text{-}Y$$

wherein, at the 5' end of the molecule, X is hydrogen or together with Y defines a covalent bond, and at the 3' end of the molecule, Y is hydrogen or a metal or together with X defines the covalent bond, each occurance of R_1 and R_3 is independently any nucleic acid residue, m is an integer between 1 and 3000 or zero , n is an integer between 1 and 3000 or zero, and R_2 is a nucleic acid sequence of the invention, particularly a nucleic acid sequence selected from Table 1. In the polynucleotide formula above R_2 is oriented so that its 5' end residue is at the left, bound to R_1, and its 3' end residue is at the right, bound to R_3. Any stretch of nucleic acid residues denoted by either R group, where m and/or n is greater than 1, may be either a heteropolymer or a homopolymer, preferably a heteropolymer. Where, in a preferred embodiment, X and Y together define a covalent bond, the polynucleotide of the above formula is a closed, circular polynucleotide, which can be a double-stranded polynucleotide wherein the formula shows a first strand to which the second strand is complementary. In another preferred embodiment m and/or n is an integer between 1 and 1000.

It is most preferred that the polynucleotides of the inventions are derived from *Staphylococcus aureus*, however, they may preferably be obtained from organisms of the same taxonomic genus. They may also be obtained, for example, from organisims of the same taxonomic family or order.

The term "polynucleotide encoding a polypeptide" as used herein encompasses polynucleotides that include a sequence encoding a polypeptide of the invention, particularly a bacterial polypeptide and more particularly a polypeptide of the *Staphylococcus aureus* Histidine Kinase having an amino acid sequence set out in Table 1 [SEQ ID NO:2]. The term also encompasses polynucleotides that include a single continuous region or discontinuous regions encoding the polypeptide (for example, interrupted by integrated phage or an insertion sequence or editing) together with additional regions, that also may contain coding and/or non-coding sequences.

The invention further relates to variants of the polynucleotides described herein that encode for variants of the polypeptide having a deduced amino acid sequence of Table 1 [SEQ ID NO:2]. Variants that are fragments of the polynucleotides of the invention may be used to synthesize full-length polynucleotides of the invention.

Further particularly preferred embodiments are polynucleotides encoding Histidine Kinase variants, that have the amino acid sequence of Histidine Kinase polypeptide of Table 1 [SEQ ID NO:2] in which several, a few, 5 to 10, 1 to 5, 1 to 3, 2, 1 or no amino acid residues are substituted, deleted or added, in any combination. Especially preferred among these are silent substitutions, additions and deletions, that do not alter the properties and activities of Histidine Kinase.

Further preferred embodiments of the invention are polynucleotides that are at least 70% identical over their entire length to a polynucleotide encoding Histidine Kinase polypeptide having an amino acid sequence set out in Table 1 [SEQ ID NO:2], and polynucleotides that are complementary to such polynucleotides. Alternatively, most highly preferred are polynucleotides that comprise a region that is at least 80% identical over its entire length to a polynucleotide encoding Histidine Kinase polypeptide and polynucleotides complementary thereto. In this regard, polynucleotides at least 90% identical over their entire length to the same are particularly preferred, and among these particularly preferred polynucleotides, those with at least 95% are especially preferred. Furthermore, those with at least 97% are highly preferred among those with at least 95%, and among these those with at least 98% and at least 99% are particularly highly preferred, with at least 99% being the more preferred.

Preferred embodiments are polynucleotides that encode polypeptides that retain substantially the same biological function or activity as the mature polypeptide encoded by a DNA of Table 1 [SEQ ID NO:1].

The invention further relates to polynucleotides that hybridize to the herein above-described sequences. In this regard, the invention especially relates to polynucleotides that hybridize under stringent conditions to the herein above-described polynucleotides. As herein used, the terms "stringent conditions" and "stringent hybridization conditions" mean hybridization will occur only if there is at least 95% and preferably at least 97% identity between the sequences. An example of stringent hybridization conditions is overnight incubation at 42°C in a solution comprising: 50% formamide, 5x SSC (150mM NaCl, 15mM trisodium citrate), 50 mM sodium

phosphate (pH7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 micrograms/ml denatured, sheared salmon sperm DNA, followed by washing the hybridization support in 0.1x SSC at about 65°C. Hybridization and wash conditions are well known and exemplified in Sambrook, *et al.*, Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor, N.Y., (1989), particularly Chapter 11 therein.

The invention also provides a polynucleotide consisting essentially of a polynucleotide sequence obtainable by screening an appropriate library containing the complete gene for a polynucleotide sequence set forth in SEQ ID NO:1 under stringent hybridization conditions with a probe having the sequence of said polynucleotide sequence set forth in SEQ ID NO:1 or a fragment thereof; and isolating said DNA sequence. Fragments useful for obtaining such a polynucleotide include, for example, probes and primers described elsewhere herein.

As discussed additionally herein regarding polynucleotide assays of the invention, for instance, polynucleotides of the invention as discussed above, may be used as a hybridization probe for RNA, cDNA and genomic DNA to isolate full-length cDNAs and genomic clones encoding Histidine Kinase and to isolate cDNA and genomic clones of other genes that have a high sequence similarity to the Histidine Kinase gene. Such probes generally will comprise at least 15 bases. Preferably, such probes will have at least 30 bases and may have at least 50 bases. Particularly preferred probes will have at least 30 bases and will have 50 bases or less.

For example, the coding region of the Histidine Kinase gene may be isolated by screening using a DNA sequence provided in Table 1 [SEQ ID NO: 1] to synthesize an oligonucleotide probe. A labeled oligonucleotide having a sequence complementary to that of a gene of the invention is then used to screen a library of cDNA, genomic DNA or mRNA to determine which members of the library the probe hybridizes to.

The polynucleotides and polypeptides of the invention may be employed, for example, as research reagents and materials for discovery of treatments of and diagnostics for disease, particularly human disease, as further discussed herein relating to polynucleotide assays.

Polynucleotides of the invention that are oligonucleotides derived from the sequences of Table 1 [SEQ ID NOS:1 or 2] may be used in the processes herein as described, but preferably for PCR, to determine whether or not the polynucleotides identified herein in whole or in part are transcribed in bacteria in infected tissue. It is recognized that such sequences will

also have utility in diagnosis of the stage of infection and type of infection the pathogen has attained.

The invention also provides polynucleotides that may encode a polypeptide that is the mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids interior to the mature polypeptide (when the mature form has more than one polypeptide chain, for instance). Such sequences may play a role in processing of a protein from precursor to a mature form, may allow protein transport, may lengthen or shorten protein half-life or may facilitate manipulation of a protein for assay or production, among other things. As generally is the case *in vivo*, the additional amino acids may be processed away from the mature protein by cellular enzymes.

A precursor protein, having the mature form of the polypeptide fused to one or more prosequences may be an inactive form of the polypeptide. When prosequences are removed such inactive precursors generally are activated. Some or all of the prosequences may be removed before activation. Generally, such precursors are called proproteins.

In addition to the standard A, G, C, T/U representations for nucleic acid bases, the term "N" may also be used in describing certain polynucleotides of the invention. "N" means that any of the four DNA or RNA bases may appear at such a designated position in the DNA or RNA sequence, except it is preferred that N is not a base that when taken in combination with adjacent nucleotide positions, when read in the correct reading frame, would have the effect of generating a premature termination codon in such reading frame.

In sum, a polynucleotide of the invention may encode a mature protein, a mature protein plus a leader sequence (which may be referred to as a preprotein), a precursor of a mature protein having one or more prosequences that are not the leader sequences of a preprotein, or a preproprotein, which is a precursor to a proprotein, having a leader sequence and one or more prosequences, which generally are removed during processing steps that produce active and mature forms of the polypeptide.

Vectors, host cells, expression

The invention also relates to vectors that comprise a polynucleotide or polynucleotides of the invention, host cells that are genetically engineered with vectors of the invention and the production of polypeptides of the invention by recombinant techniques. Cell-free translation

systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the invention.

For recombinant production, host cells can be genetically engineered to incorporate expression systems or portions thereof or polynucleotides of the invention. Introduction of a polynucleotide into the host cell can be effected by methods described in many standard laboratory manuals, such as Davis et al., *BASIC METHODS IN MOLECULAR BIOLOGY*, (1986) and Sambrook et al., *MOLECULAR CLONING: A LABORATORY MANUAL*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989), such as, calcium phosphate transfection, DEAE-dextran mediated transfection, transvection, microinjection, cationic lipid-mediated transfection, electroporation, transduction, scrape loading, ballistic introduction and infection.

Representative examples of appropriate hosts include bacterial cells, such as streptococci, staphylococci, enterococci *E. coli*, streptomyces and *Bacillus subtilis* cells; fungal cells, such as yeast cells and *Aspergillus* cells; insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS, HeLa, C127, 3T3, BHK, 293 and Bowes melanoma cells; and plant cells.

A great variety of expression systems can be used to produce the polypeptides of the invention. Such vectors include, among others, chromosomal, episomal and virus-derived vectors, *e.g.*, vectors derived from bacterial plasmids, from bacteriophage, from transposons, from yeast episomes, from insertion elements, from yeast chromosomal elements, from viruses such as baculoviruses, papova viruses, such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, and vectors derived from combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids. The expression system constructs may contain control regions that regulate as well as engender expression. Generally, any system or vector suitable to maintain, propagate or express polynucleotides and/or to express a polypeptide in a host may be used for expression in this regard. The appropriate DNA sequence may be inserted into the expression system by any of a variety of well-known and routine techniques, such as, for example, those set forth in Sambrook *et al.*, *MOLECULAR CLONING, A LABORATORY MANUAL*, (*supra*).

For secretion of the translated protein into the lumen of the endoplasmic reticulum, into the periplasmic space or into the extracellular environment, appropriate secretion signals may be

incorporated into the expressed polypeptide. These signals may be endogenous to the polypeptide or they may be heterologous signals.

Polypeptides of the invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography, and lectin chromatography. Most preferably, high performance liquid chromatography is employed for purification. Well known techniques for refolding protein may be employed to regenerate active conformation when the polypeptide is denatured during isolation and or purification.

Diagnostic Assays

This invention is also related to the use of the Histidine Kinase polynucleotides of the invention for use as diagnostic reagents. Detection of Histidine Kinase in a eukaryote, particularly a mammal, and especially a human, will provide a diagnostic method for diagnosis of a disease. Eukaryotes (herein also "individual(s)"), particularly mammals, and especially humans, particularly those infected or suspected to be infected with an organism comprising the Histidine Kinase gene may be detected at the nucleic acid level by a variety of techniques.

Nucleic acids for diagnosis may be obtained from an infected individual's cells and tissues, such as bone, blood, muscle, cartilage, and skin. Genomic DNA may be used directly for detection or may be amplified enzymatically by using PCR or other amplification technique prior to analysis. RNA, cDNA and genomic DNA may also be used in the same ways. Using amplification, characterization of the species and strain of prokaryote present in an individual, may be made by an analysis of the genotype of the prokaryote gene. Deletions and insertions can be detected by a change in size of the amplified product in comparison to the genotype of a reference sequence. Point mutations can be identified by hybridizing amplified DNA to labeled Histidine Kinase polynucleotide sequences. Perfectly matched sequences can be distinguished from mismatched duplexes by RNase digestion or by differences in melting temperatures. DNA sequence differences may also be detected by alterations in the electrophoretic mobility of the DNA fragments in gels, with or without denaturing agents, or by direct DNA sequencing. See, *e.g.*, Myers et al., *Science, 230:* 1242 (1985). Sequence changes at specific locations also may be

revealed by nuclease protection assays, such as RNase and S1 protection or a chemical cleavage method. See, *e.g.*, Cotton et al., *Proc. Natl. Acad. Sci., USA, 85:* 4397-4401 (1985).

Cells carrying mutations or polymorphisms in the gene of the invention may also be detected at the DNA level by a variety of techniques, to allow for serotyping, for example. For example, RT-PCR can be used to detect mutations. It is particularly preferred to used RT-PCR in conjunction with automated detection systems, such as, for example, GeneScan. RNA, cDNA or genomic DNA may also be used for the same purpose, PCR or RT-PCR. As an example, PCR primers complementary to a nucleic acid encoding Histidine Kinase can be used to identify and analyze mutations. Examples of representative primers are shown below in Table 2.

Table 2

Primers for amplification of Histidine Kinase polynucleotides

SEQ ID NO	PRIMER SEQUENCE
5	5'-ATGACAAAACGTAAATTGCGCAATAAC-3'
6	5'-AAATATGATTTTAAACGTTGTTCC-3'

The invention also includes primers of the formula:

$$X\text{-}(R_1)_m\text{-}(R_2)\text{-}(R_3)_n\text{-}Y$$

wherein, at the 5' end of the molecule, X is hydrogen, and at the 3' end of the molecule, Y is hydrogen or a metal, R_1 and R_3 is any nucleic acid residue, m is an integer between 1 and 20 or zero , n is an integer between 1 and 20 or zero, and R_2 is a primer sequence of the invention, particularly a primer sequence selected from Table 2. In the polynucleotide formula above R_2 is oriented so that its 5' end residue is at the left, bound to R_1, and its 3' end residue is at the right, bound to R_3. Any stretch of nucleic acid residues denoted by either R group, where m and/or n is greater than 1, may be either a heteropolymer or a homopolymer, preferably a heteropolymer being complementary to a region of a polynucleotide of Table 1. In a preferred embodiment m and/or n is an integer between 1 and 10.

The invention further provides these primers with 1, 2, 3 or 4 nucleotides removed from the 5' and/or the 3' end. These primers may be used for, among other things, amplifying Histidine Kinase DNA isolated from a sample derived from an individual. The primers may be used to

amplify the gene isolated from an infected individual such that the gene may then be subject to various techniques for elucidation of the DNA sequence. In this way, mutations in the DNA sequence may be detected and used to diagnose infection and to serotype and/or classify the infectious agent.

The invention further provides a process for diagnosing, disease, preferably bacterial infections, more preferably infections by *Staphylococcus aureus*, comprising determining from a sample derived from an individual a increased level of expression of polynucleotide having a sequence of Table 1 [SEQ ID NO: 1]. Increased or decreased expression of Histidine Kinase polynucleotide can be measured using any on of the methods well known in the art for the quantation of polynucleotides, such as, for example, amplification, PCR, RT-PCR, RNase protection, Northern blotting and other hybridization methods.

In addition, a diagnostic assay in accordance with the invention for detecting over-expression of Histidine Kinase protein compared to normal control tissue samples may be used to detect the presence of an infection, for example. Assay techniques that can be used to determine levels of a Histidine Kinase protein, in a sample derived from a host are well-known to those of skill in the art. Such assay methods include radioimmunoassays, competitive-binding assays, Western Blot analysis and ELISA assays.

Antibodies

The polypeptides of the invention or variants thereof, or cells expressing them can be used as an immunogen to produce antibodies immunospecific for such polypeptides. "Antibodies" as used herein includes monoclonal and polyclonal antibodies, chimeric, single chain, simianized antibodies and humanized antibodies, as well as Fab fragments, including the products of an Fab immunoglobulin expression library.

Antibodies generated against the polypeptides of the invention can be obtained by administering the polypeptides or epitope-bearing fragments, analogues or cells to an animal, preferably a nonhuman, using routine protocols. For preparation of monoclonal antibodies, any technique known in the art that provides antibodies produced by continuous cell line cultures can be used. Examples include various techniques, such as those in Kohler, G. and Milstein, C., *Nature 256:* 495-497 (1975); Kozbor *et al.*, *Immunology Today 4:* 72 (1983); Cole et al., pg. 77-96 in *MONOCLONAL ANTIBODIES AND CANCER THERAPY*, Alan R. Liss, Inc. (1985).

Techniques for the production of single chain antibodies (U.S. Patent No. 4,946,778) can be adapted to produce single chain antibodies to polypeptides of this invention. Also, transgenic mice, or other organisms such as other mammals, may be used to express humanized antibodies.

Alternatively phage display technology may be utilized to select antibody genes with binding activities towards the polypeptide either from repertoires of PCR amplified v-genes of lymphocytes from humans screened for possessing anti-Histidine Kinase or from naive libraries (McCafferty, J. et al., (1990), Nature 348, 552-554; Marks, J. et al., (1992) Biotechnology *10*, 779-783). The affinity of these antibodies can also be improved by chain shuffling (Clackson, T. et al., (1991) Nature 352, 624-628).

If two antigen binding domains are present each domain may be directed against a different epitope - termed 'bispecific' antibodies.

The above-described antibodies may be employed to isolate or to identify clones expressing the polypeptides to purify the polypeptides by affinity chromatography.

Thus, among others, antibodies against Histidine Kinase-polypeptide may be employed to treat infections, particularly bacterial infections.

Polypeptide variants include antigenically, epitopically or immunologically equivalent variants that form a particular aspect of this invention. The term "antigenically equivalent derivative" as used herein encompasses a polypeptide or its equivalent which will be specifically recognized by certain antibodies which, when raised to the protein or polypeptide according to the invention, interfere with the immediate physical interaction between pathogen and mammalian host. The term "immunologically equivalent derivative" as used herein encompasses a peptide or its equivalent which when used in a suitable formulation to raise antibodies in a vertebrate, the antibodies act to interfere with the immediate physical interaction between pathogen and mammalian host.

The polypeptide, such as an antigenically or immunologically equivalent derivative or a fusion protein thereof is used as an antigen to immunize a mouse or other animal such as a rat or chicken. The fusion protein may provide stability to the polypeptide. The antigen may be associated, for example by conjugation, with an immunogenic carrier protein for example bovine serum albumin (BSA) or keyhole limpet haemocyanin (KLH). Alternatively a multiple antigenic peptide comprising multiple copies of the protein or polypeptide, or an antigenically

or immunologically equivalent polypeptide thereof may be sufficiently antigenic to improve immunogenicity so as to obviate the use of a carrier.

Preferably, the antibody or variant thereof is modified to make it less immunogenic in the individual. For example, if the individual is human the antibody may most preferably be "humanized"; where the complimentarity determining region(s) of the hybridoma-derived antibody has been transplanted into a human monoclonal antibody , for example as described in Jones, P. *et al.* (1986), *Nature* 321, 522-525 or Tempest *et al.*, (1991) *Biotechnology* 9, 266-273.

The use of a polynucleotide of the invention in genetic immunization will preferably employ a suitable delivery method such as direct injection of plasmid DNA into muscles (Wolff et al., Hum Mol Genet 1992, 1:363, Manthorpe et al., Hum. Gene Ther. 1963:4, 419), delivery of DNA complexed with specific protein carriers (Wu et al., *J Biol Chem.* 1989: 264,16985), coprecipitation of DNA with calcium phosphate (Benvenisty & Reshef, *PNAS USA*, 1986:83,9551), encapsulation of DNA in various forms of liposomes (Kaneda et al., *Science* 1989:243,375), particle bombardment (Tang et al., *Nature* 1992, 356:152, Eisenbraun et al., *DNA Cell Biol* 1993, 12:791) and *in vivo* infection using cloned retroviral vectors (Seeger et al., *PNAS USA* 1984:81,5849).

Antagonists and agonists - assays and molecules

Polypeptides of the invention may also be used to assess the binding of small molecule substrates and ligands in, for example, cells, cell-free preparations, chemical libraries, and natural product mixtures. These substrates and ligands may be natural substrates and ligands or may be structural or functional mimetics. See, *e.g.*, Coligan *et al.*, *Current Protocols in Immunology 1(2):* Chapter 5 (1991).

The invention also provides a method of screening compounds to identify those which enhance (agonist) or block (antagonist) the action of Histidine Kinase polypeptides or polynucleotides, particularly those compounds that are bacteriostatic and/or bacteriocidal. The method of screening may involve high-throughput techniques. For example, to screen for agonists or antagoists, a synthetic reaction mix, a cellular compartment, such as a membrane, cell envelope or cell wall, or a preparation of any thereof, comprising Histidine Kinase polypeptide and a labeled substrate or ligand of such polypeptide is incubated in the absence or the presence of a candidate molecule that may be a Histidine Kinase agonist or antagonist. The ability of the

candidate molecule to agonize or antagonize the Histidine Kinase polypeptide is reflected in decreased binding of the labeled ligand or decreased production of product from such substrate. Molecules that bind gratuitously, *i.e.*, without inducing the effects of Histidine Kinase polypeptide are most likely to be good antagonists. Molecules that bind well and increase the rate of product production from substrate are agonists. Detection of the rate or level of production of product from substrate may be enhanced by using a reporter system. Reporter systems that may be useful in this regard include but are not limited to colorimetric labeled substrate converted into product, a reporter gene that is responsive to changes in Histidine Kinase polynucleotide or polypeptide activity, and binding assays known in the art.

Another example of an assay for Histidine Kinase antagonists is a competitive assay that combines Histidine Kinase and a potential antagonist with Histidine Kinase-binding molecules, recombinant Histidine Kinase binding molecules, natural substrates or ligands, or substrate or ligand mimetics, under appropriate conditions for a competitive inhibition assay. Histidine Kinase can be labeled, such as by radioactivity or a colorimetric compound, such that the number of Histidine Kinase molecules bound to a binding molecule or converted to product can be determined accurately to assess the effectiveness of the potential antagonist.

Potential antagonists include small organic molecules, peptides, polypeptides and antibodies that bind to a polynucleotide or polypeptide of the invention and thereby inhibit or extinguish its activity. Potential antagonists also may be small organic molecules, a peptide, a polypeptide such as a closely related protein or antibody that binds the same sites on a binding molecule, such as a binding molecule, without inducing Histidine Kinase-induced activities, thereby preventing the action of Histidine Kinase by excluding Histidine Kinase from binding.

Potential antagonists include a small molecule that binds to and occupies the binding site of the polypeptide thereby preventing binding to cellular binding molecules, such that normal biological activity is prevented. Examples of small molecules include but are not limited to small organic molecules, peptides or peptide-like molecules. Other potential antagonists include antisense molecules (see Okano, *J. Neurochem. 56:* 560 (1991); *OLIGODEOXYNUCLEOTIDES AS ANTISENSE INHIBITORS OF GENE EXPRESSION*, CRC Press, Boca Raton, FL (1988), for a description of these molecules). Preferred potential antagonists include compounds related to and variants of Histidine Kinase.

Each of the DNA sequences provided herein may be used in the discovery and development of antibacterial compounds. The encoded protein, upon expression, can be used as a target for the screening of antibacterial drugs. Additionally, the DNA sequences encoding the amino terminal regions of the encoded protein or Shine-Delgarno or other translation facilitating sequences of the respective mRNA can be used to construct antisense sequences to control the expression of the coding sequence of interest.

The invention also provides the use of the polypeptide, polynucleotide or inhibitor of the invention to interfere with the initial physical interaction between a pathogen and mammalian host responsible for sequelae of infection. In particular the molecules of the invention may be used: in the prevention of adhesion of bacteria, in particular gram positive bacteria, to mammalian extracellular matrix proteins on in-dwelling devices or to extracellular matrix proteins in wounds; to block Histidine Kinase protein-mediated mammalian cell invasion by, for example, initiating phosphorylation of mammalian tyrosine kinases (Rosenshine *et al.*, *Infect. Immun. 60*:2211 (1992); to block bacterial adhesion between mammalian extracellular matrix proteins and bacterial Histidine Kinase proteins that mediate tissue damage and; to block the normal progression of pathogenesis in infections initiated other than by the implantation of in-dwelling devices or by other surgical techniques.

This invention provides a method of screening drugs to identify those which interfere with i) the interaction of the histidine kinase with a response regulator, the method comprising incubating the histidine kinase with response regulator in the presence of the drug and measuring the ability of the drug to block this interaction; and/or ii) the ability of the histidine kinase to autophosphorylate, the method comprising incubating the histidine kinase with the drug and measuring the ability of the drug to prevent autophosphorylation.

The response regulator is preferably the cognate response regulator of the histidine kinase, or another response regulator which is capable of using the histidine kinase as a substrate, and is preferably from *Staphylococcus aureus* or another microorganism e.g. *Bacillus*. Polypeptide and polynucleotide sequences of the cognate response regulator of the Histidine kinase of the invention are set forth in Table 1 (C and D). This novel response regulator shows 42% identity to the ResD response regulator protein from *Bacillus subtilis*.

The antagonists and agonists of the invention may be employed, for instance, to inhibit and treat diseases.

Helicobacter pylori (herein *H. pylori*) bacteria infect the stomachs of over one-third of the world's population causing stomach cancer, ulcers, and gastritis (International Agency for Research on Cancer (1994) Schistosomes, Liver Flukes and Helicobacter Pylori (International Agency for Research on Cancer, Lyon, France; http://www.uicc.ch/ecp/ecp2904.htm). Moreover, the international Agency for Research on Cancer recently recognized a cause-and-effect relationship between *H. pylori* and gastric adenocarcinoma, classifying the bacterium as a Group I (definite) carcinogen. Preferred antimicrobial compounds of the invention (agonists and antagonists of Histidine Kinase) found using screens provided by the invention, particularly broad-spectrum antibiotics, should be useful in the treatment of *H. pylori* infection. Such treatment should decrease the advent of *H. pylori*-induced cancers, such as gastrointestinal carcinoma. Such treatment should also cure gastric ulcers and gastritis.

Vaccines

Another aspect of the invention relates to a method for inducing an immunological response in an individual, particularly a mammal which comprises inoculating the individual with Histidine Kinase, or a fragment or variant thereof, adequate to produce antibody and/ or T cell immune response to protect said individual from infection, particularly bacterial infection and most particularly *Staphylococcus aureus* infection. Also provided are methods whereby such immunological response slows bacterial replication. Yet another aspect of the invention relates to a method of inducing immunological response in an individual which comprises delivering to such individual a nucleic acid vector to direct expression of Histidine Kinase, or a fragment or a variant thereof, for expressing Histidine Kinase, or a fragment or a variant thereof *in vivo* in order to induce an immunological response, such as, to produce antibody and/ or T cell immune response, including, for example, cytokine-producing T cells or cytotoxic T cells, to protect said individual from disease, whether that disease is already established within the individual or not. One way of administering the gene is by accelerating it into the desired cells as a coating on particles or otherwise. Such nucleic acid vector may comprise DNA, RNA, a modified nucleic acid, or a DNA/RNA hybrid.

A further aspect of the invention relates to an immunological composition which, when introduced into an individual capable or having induced within it an immunological response, induces an immunological response in such individual to a Histidine Kinase or protein coded therefrom, wherein the composition comprises a recombinant Histidine Kinase or protein coded

therefrom comprising DNA which codes for and expresses an antigen of said Histidine Kinase or protein coded therefrom. The immunological response may be used therapeutically or prophylactically and may take the form of antibody immunity or cellular immunity such as that arising from CTL or CD4+ T cells.

A Histidine Kinase polypeptide or a fragment thereof may be fused with co-protein which may not by itself produce antibodies, but is capable of stabilizing the first protein and producing a fused protein which will have immunogenic and protective properties. Thus fused recombinant protein, preferably further comprises an antigenic co-protein, such as lipoprotein D from *Hemophilus influenzae*, Glutathione-S-transferase (GST) or beta-galactosidase, relatively large co-proteins which solubilize the protein and facilitate production and purification thereof. Moreover, the co-protein may act as an adjuvant in the sense of providing a generalized stimulation of the immune system. The co-protein may be attached to either the amino or carboxy terminus of the first protein.

Provided by this invention are compositions, particularly vaccine compositions, and methods comprising the polypeptides or polynucleotides of the invention and immunostimulatory DNA sequences, such as those described in Sato, Y. *et al.* Science 273: 352 (1996).

Also, provided by this invention are methods using the described polynucleotide or particular fragments thereof which have been shown to encode non-variable regions of bacterial cell surface proteins in DNA constructs used in such genetic immunization experiments in animal models of infection with *Staphylococcus aureus* will be particularly useful for identifying protein epitopes able to provoke a prophylactic or therapeutic immune response. It is believed that this approach will allow for the subsequent preparation of monoclonal antibodies of particular value from the requisite organ of the animal successfully resisting or clearing infection for the development of prophylactic agents or therapeutic treatments of bacterial infection, particularly *Staphylococcus aureus* infection, in mammals, particularly humans.

The polypeptide may be used as an antigen for vaccination of a host to produce specific antibodies which protect against invasion of bacteria, for example by blocking adherence of bacteria to damaged tissue. Examples of tissue damage include wounds in skin or connective tissue caused, e.g., by mechanical, chemical or thermal damage or by implantation of

indwelling devices, or wounds in the mucous membranes, such as the mouth, mammary glands, urethra or vagina.

The invention also includes a vaccine formulation which comprises an immunogenic recombinant protein of the invention together with a suitable carrier. Since the protein may be broken down in the stomach, it is preferably administered parenterally, including, for example, administration that is subcutaneous, intramuscular, intravenous, or intradermal. Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which may contain anti-oxidants, buffers, bacteriostats and solutes which render the formulation insotonic with the bodily fluid, preferably the blood, of the individual; and aqueous and non-aqueous sterile suspensions which may include suspending agents or thickening agents. The formulations may be presented in unit-dose or multi-dose containers, for example, sealed ampules and vials and may be stored in a freeze-dried condition requiring only the addition of the sterile liquid carrier immediately prior to use. The vaccine formulation may also include adjuvant systems for enhancing the immunogenicity of the formulation, such as oil-in water systems and other systems known in the art. The dosage will depend on the specific activity of the vaccine and can be readily determined by routine experimentation.

While the invention has been described with reference to certain Histidine Kinase protein, it is to be understood that this covers fragments of the naturally occurring protein and similar proteins with additions, deletions or substitutions which do not substantially affect the immunogenic properties of the recombinant protein.

Compositions, kits and administration

The invention also relates to compositions comprising the polynucleotide or the polypeptides discussed above or their agonists or antagonists. The polypeptides of the invention may be employed in combination with a non-sterile or sterile carrier or carriers for use with cells, tissues or organisms, such as a pharmaceutical carrier suitable for administration to a subject. Such compositions comprise, for instance, a media additive or a therapeutically effective amount of a polypeptide of the invention and a pharmaceutically acceptable carrier or excipient. Such carriers may include, but are not limited to, saline, buffered saline, dextrose, water, glycerol, ethanol and combinations thereof. The formulation should suit the mode of administration. The invention further relates to diagnostic and pharmaceutical packs and kits comprising one or more

containers filled with one or more of the ingredients of the aforementioned compositions of the invention.

Polypeptides and other compounds of the invention may be employed alone or in conjunction with other compounds, such as therapeutic compounds.

The pharmaceutical compositions may be administered in any effective, convenient manner including, for instance, administration by topical, oral, anal, vaginal, intravenous, intraperitoneal, intramuscular, subcutaneous, intranasal or intradermal routes among others.

In therapy or as a prophylactic, the active agent may be administered to an individual as an injectable composition, for example as a sterile aqueous dispersion, preferably isotonic.

Alternatively the composition may be formulated for topical application for example in the form of ointments, creams, lotions, eye ointments, eye drops, ear drops, mouthwash, impregnated dressings and sutures and aerosols, and may contain appropriate conventional additives, including, for example, preservatives, solvents to assist drug penetration, and emollients in ointments and creams. Such topical formulations may also contain compatible conventional carriers, for example cream or ointment bases, and ethanol or oleyl alcohol for lotions. Such carriers may constitute from about 1% to about 98% by weight of the formulation; more usually they will constitute up to about 80% by weight of the formulation.

For administration to mammals, and particularly humans, it is expected that the daily dosage level of the active agent will be from 0.01 mg/kg to 10 mg/kg, typically around 1 mg/kg. The physician in any event will determine the actual dosage which will be most suitable for an individual and will vary with the age, weight and response of the particular individual. The above dosages are exemplary of the average case. There can, of course, be individual instances where higher or lower dosage ranges are merited, and such are within the scope of this invention.

In-dwelling devices include surgical implants, prosthetic devices and catheters, i.e., devices that are introduced to the body of an individual and remain in position for an extended time. Such devices include, for example, artificial joints, heart valves, pacemakers, vascular grafts, vascular catheters, cerebrospinal fluid shunts, urinary catheters, continuous ambulatory peritoneal dialysis (CAPD) catheters.

The composition of the invention may be administered by injection to achieve a systemic effect against relevant bacteria shortly before insertion of an in-dwelling device. Treatment may be continued after surgery during the in-body time of the device. In addition, the composition could also be used to broaden perioperative cover for any surgical technique to prevent bacterial wound infections, especially *Staphylococcus aureus* wound infections.

Many orthopaedic surgeons consider that humans with prosthetic joints should be considered for antibiotic prophylaxis before dental treatment that could produce a bacteremia. Late deep infection is a serious complication sometimes leading to loss of the prosthetic joint and is accompanied by significant morbidity and mortality. It may therefore be possible to extend the use of the active agent as a replacement for prophylactic antibiotics in this situation.

In addition to the therapy described above, the compositions of this invention may be used generally as a wound treatment agent to prevent adhesion of bacteria to matrix proteins exposed in wound tissue and for prophylactic use in dental treatment as an alternative to, or in conjunction with, antibiotic prophylaxis.

Alternatively, the composition of the invention may be used to bathe an indwelling device immediately before insertion. The active agent will preferably be present at a concentration of 1μg/ml to 10mg/ml for bathing of wounds or indwelling devices.

A vaccine composition is conveniently in injectable form. Conventional adjuvants may be employed to enhance the immune response. A suitable unit dose for vaccination is 0.5-5 microgram/kg of antigen, and such dose is preferably administered 1-3 times and with an interval of 1-3 weeks. With the indicated dose range, no adverse toxicological effects will be observed with the compounds of the invention which would preclude their administration to suitable individuals.

Each reference disclosed herein is incorporated by reference herein in its entirety. Any patent application to which this application claims priority is also incorporated by reference herein in its entirety.

GLOSSARY

The following definitions are provided to facilitate understanding of certain terms used frequently herein.

"Disease(s)" means and disease caused by or related to infection by a bacteria, including disease, such as, infections of the upper respiratory tract (e.g., otitis media, bacterial tracheitis,

acute epiglottitis, thyroiditis), lower respiratory (e.g., empyema, lung abscess), cardiac (e.g., infective endocarditis), gastrointestinal (e.g., secretory diarrhoea, splenic absces, retroperitoneal abscess), CNS (e.g., cerebral abscess), eye (e.g., blepharitis, conjunctivitis, keratitis, endophthalmitis, preseptal and orbital cellulitis, darcryocystitis), kidney and urinary tract (e.g., epididymitis, intrarenal and perinephric absces, toxic shock syndrome), skin (e.g., impetigo, folliculitis, cutaneous abscesses, cellulitis, wound infection, bacterial myositis) bone and joint (e.g., septic arthritis, osteomyelitis).

"Host cell" is a cell which has been transformed or transfected, or is capable of transformation or transfection by an exogenous polynucleotide sequence.

"Identity," as known in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as the case may be, as determined by the match between strings of such sequences. "Identity" and "similarity" can be readily calculated by known methods, including but not limited to those described in (*Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data*, Part I, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991; and Carillo, H., and Lipman, D., SIAM *J. Applied Math., 48:* 1073 (1988). Preferred methods to determine identity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity are codified in publicly available computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited to, the GCG program package (Devereux, J., et al., *Nucleic Acids Research 12(1):* 387 (1984)), BLASTP, BLASTN, and FASTA (Atschul, S.F. et al., *J. Molec. Biol. 215:* 403-410 (1990). The BLAST X program is publicly available from NCBI and other sources (*BLAST Manual*, Altschul, S., *et al.*, NCBI NLM NIH Bethesda, MD 20894; Altschul, S., *et al., J. Mol. Biol. 215*: 403-410 (1990). The well known Smith Waterman algorithm may also be used to determine identity. Preferred embodiments include an isolated polynucleotide comprising a polynucleotide having at least a 50,60, 70, 80, 85, 90,

95, 97 or 100% identity to a polynucleotide reference sequence of SEQ ID NO:1, wherein said reference sequence may be identical to the sequence of SEQ ID NO: 1 or may include up to a certain integer number of nucleotide alterations as compared to the reference sequence, wherein said alterations are selected from the group consisting of at least one nucleotide deletion, substitution or insertion, and wherein said alterations may occur at the 5' or 3' terminal positions of the reference nucleotide sequence or anywhere between those terminal positions, interspersed either individually among the nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence, and wherein said number of nucleotide alterations is determined by multiplying the total number of nucleotides in SEQ ID NO:1 by the numerical percent of the respective percent identity and subtracting that product from said total number of nucleotides in SEQ ID NO:1, or:

$$n \leq x - (x \bullet y),$$

wherein n is the number of nucleotide alterations, x is the total number of nucleotides in SEQ ID NO:1, and y is 0.50 for 50%, 0.60 for 60%, 0.70 for 70%, 0.80 for 80%, 0.85 for 85%, 0.90 for 90%, 0.95 for 95%, 0.97 for 97% or 1.00 for 100%, and wherein any non-integer product of x and y is rounded down to the nearest integer prior to subtracting it from x.

"Isolated" means altered "by the hand of man" from its natural state, *i.e.*, if it occurs in nature, it has been changed or removed from its original environment, or both. For example, a polynucleotide or a polypeptide naturally present in a living organism is not "isolated," but the same polynucleotide or polypeptide separated from the coexisting materials of its natural state is "isolated", as the term is employed herein. Moreover, a polynucleotide or polypeptide that is introduced into an organism by transformation, genetic manipulation or by any other recombinant method is "isolated" even if it is still present in said organism, which organism may be living or non-living.

"Polynucleotide(s)" generally refers to any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. "Polynucleotide(s)" include, without limitation, single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions or single-, double- and triple-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules

comprising DNA and RNA that may be single-stranded or, more typically, double-stranded, or triple-stranded regions, or a mixture of single- and double-stranded regions. In addition, "polynucleotide" as used herein refers to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The strands in such regions may be from the same molecule or from different molecules. The regions may include all of one or more of the molecules, but more typically involve only a region of some of the molecules. One of the molecules of a triple-helical region often is an oligonucleotide. As used herein, the term "polynucleotide(s)" also includes DNAs or RNAs as described above that contain one or more modified bases. Thus, DNAs or RNAs with backbones modified for stability or for other reasons are "polynucleotide(s)" as that term is intended herein. Moreover, DNAs or RNAs comprising unusual bases, such as inosine, or modified bases, such as tritylated bases, to name just two examples, are polynucleotides as the term is used herein. It will be appreciated that a great variety of modifications have been made to DNA and RNA that serve many useful purposes known to those of skill in the art. The term "polynucleotide(s)" as it is employed herein embraces such chemically, enzymatically or metabolically modified forms of polynucleotides, as well as the chemical forms of DNA and RNA characteristic of viruses and cells, including, for example, simple and complex cells. "Polynucleotide(s)" also embraces short polynucleotides often referred to as oligonucleotide(s).

"Polypeptide(s)" refers to any peptide or protein comprising two or more amino acids joined to each other by peptide bonds or modified peptide bonds. "Polypeptide(s)" refers to both short chains, commonly referred to as peptides, oligopeptides and oligomers and to longer chains generally referred to as proteins. Polypeptides may contain amino acids other than the 20 gene encoded amino acids. "Polypeptide(s)" include those modified either by natural processes, such as processing and other post-translational modifications, but also by chemical modification techniques. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature, and they are well known to those of skill in the art. It will be appreciated that the same type of modification may be present in the same or varying degree at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains, and the amino or carboxyl termini. Modifications include, for example, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide

or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins, such as arginylation, and ubiquitination. See, for instance, *PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES*, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York (1993) and Wold, F., Posttranslational Protein Modifications: Perspectives and Prospects, pgs. 1-12 in *POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS*, B. C. Johnson, Ed., Academic Press, New York (1983); Seifter et al., *Meth. Enzymol. 182*:626-646 (1990) and Rattan et al., *Protein Synthesis: Posttranslational Modifications and Aging*, Ann. N.Y. Acad. Sci. 663: 48-62 (1992). Polypeptides may be branched or cyclic, with or without branching. Cyclic, branched and branched circular polypeptides may result from post-translational natural processes and may be made by entirely synthetic methods, as well.

"Variant(s)" as the term is used herein, is a polynucleotide or polypeptide that differs from a reference polynucleotide or polypeptide respectively, but retains essential properties. A typical variant of a polynucleotide differs in nucleotide sequence from another, reference polynucleotide. Changes in the nucleotide sequence of the variant may or may not alter the amino acid sequence of a polypeptide encoded by the reference polynucleotide. Nucleotide changes may result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence, as discussed below. A typical variant of a polypeptide differs in amino acid sequence from another, reference polypeptide. Generally, differences are limited so that the sequences of the reference polypeptide and the variant are closely similar overall and, in many regions, identical. A variant and reference polypeptide may differ in amino acid sequence by one or more substitutions, additions, deletions in any combination. A substituted or inserted amino acid residue may or may not be one encoded by the genetic code. A variant of a polynucleotide or polypeptide may be a naturally occurring such as an allelic variant, or it may be a variant that is not known to occur naturally. Non-

naturally occurring variants of polynucleotides and polypeptides may be made by mutagenesis techniques, by direct synthesis, and by other recombinant methods known to skilled artisans.

EXAMPLES

The examples below are carried out using standard techniques, which are well known and routine to those of skill in the art, except where otherwise described in detail. The examples are illustrative, but do not limit the invention.

Example 1 Strain selection, Library Production and Sequencing

The polynucleotide having a DNA sequence given in Table 1 [SEQ ID NO:1] was obtained from a library of clones of chromosomal DNA of *Staphylococcus aureus* in *E. coli*. The sequencing data from two or more clones containing overlapping *Staphylococcus aureus* DNAs was used to construct the contiguous DNA sequence in SEQ ID NO:1. Libraries may be prepared by routine methods, for example:

Methods 1 and 2 below.

Total cellular DNA is isolated from *Staphylococcus aureus* WCUH 29 according to standard procedures and size-fractionated by either of two methods.

Method 1

Total cellular DNA is mechanically sheared by passage through a needle in order to size-fractionate according to standard procedures. DNA fragments of up to 11kbp in size are rendered blunt by treatment with exonuclease and DNA polymerase, and EcoRI linkers added. Fragments are ligated into the vector Lambda ZapII that has been cut with EcoRI, the library packaged by standard procedures and *E.coli* infected with the packaged library. The library is amplified by standard procedures.

Method 2

Total cellular DNA is partially hydrolyzed with a one or a combination of restriction enzymes appropriate to generate a series of fragments for cloning into library vectors (e.g., RsaI, PalI, AluI, Bshl235I), and such fragments are size-fractionated according to standard procedures. EcoRI linkers are ligated to the DNA and the fragments then ligated into the vector Lambda ZapII that have been cut with EcoRI, the library packaged by standard procedures, and *E.coli* infected with the packaged library. The library is amplified by standard procedures.

SEQUENCE LISTING

(1) GENERAL INFORMATION

(i) APPLICANT: Wallis, Nicola G.
Shilling, Lisa K.
Mooney, Jeffrey L.
Debouck, Christine
Zhong, YiYi
Jaworski, Deborah D.
Wang, Min
Throup, John P.

(ii) TITLE OF THE INVENTION: Histidine Kinase

(iii) NUMBER OF SEQUENCES: 6

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Dechert, Price & Rhoads
(B) STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
(C) CITY: Philadelphia
(D) STATE: PA
(E) COUNTRY: USA
(F) ZIP: 19103-2793

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Diskette
(B) COMPUTER: IBM Compatible
(C) OPERATING SYSTEM: DOS
(D) SOFTWARE: FastSEQ for Windows Version 2.0

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER:
(B) FILING DATE:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Falk, Stephen T
(B) REGISTRATION NUMBER: 36,795
(C) REFERENCE/DOCKET NUMBER: GM10127

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 215-994-2488
(B) TELEFAX: 215-994-2222
(C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2201 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

TAATTTAAAA AGCAACTATT GTATAGAAAA ATACAAAATT TAAAATATAT TACCTTATTA
60
GAAAAAGTCG TAATATGAGG TGTACAAATG ACGCAAATTT TAATAGTAGA AGATGAACAA
120
AACTTAGCAA GATTTCTTGA ATTGGAACTC ACACATGAAA ATTACAATGT GGACACAGAG
180
TATGATGGAC AAGACGGTTT AGATAAAGCG CTTAGCCATT ACTATGATTT AATCATATTA
240
GATTTAATGT TGCCGTCAAT TAATGGCTTA GAAATTTGTC GCAAAATTAG ACAACAACAA
300
TCTACACCTA TCATTATAAT TACAGCGAAA AGTGATACGT ATGACAAAGT TGCTGGGCTT
360

GATTACGGTG CAGACGATTA TATAGTTAAG CCGTTTGATA TTGAAGAACT TTTAGCAAGA
420

ATTCGTGCAA TTTTACGTCG TCAGCCACAA AAGGATATTA TCGATGTCAA CGGTATTACA
480

ATTGATAAGA ACGCTTTTAA AGTGACGGTA AATGGCGCAG AAATTGAATT AACAAAAACA
540

GAGTATGATT TACTATATCT TCTAGCTGAA AATAAAAACC ATGTTATGCA ACGGGAACAA
600

ATTTTAAATC ATGTATGGGG TTATAATAGT GAAGTAGAAA CAAATGTCGT AGATGTTTAT
660

ATAAGATATT TACGAAACAA GTTAAAACCA TACGATCGTG ACAAAATGAT TGAAACAGTT
720

CGTGGCGTTG GGTATGTGAT ACGATGACAA AACGTAAATT GCGCAATAAC TGGATTATTG
780

TTACCACGAT GATTACGTTT GTCACGATAT TTTTGTTTTG TTTAATTATT ATTTTTTTCT
840

TGAAAGATAC ACTGCATAAT AGTGAGCTTG ATGATGCAGA ACGAAGCTCA AGCGATATTA
900

ATAATTTATT TCATTCTAAG CCTGTTAAAG ATATATCTGC ATTAGACTTG AATGCATCTT
960

TAGGTAATTT TCAAGAGATA ATTATTTATG ATGAGCATAA TAATAAATTA TTTGAGACAT
1020

CGAATGATAA CACAGTGAGA GTTGAACCAG GTTATGAACA CCGTTATTTT GACCGCGTAA
1080

TAAAAAAACG CTATAAAGGC ATTGAATATT TAATTATTAA AGAACCAATT ACAACGCAAG
1140

ATTTCAAAGG GTATAGCTTG TTAATTCATT CACTAGAAAA TTATGATAAC ATCGTAAAAT
1200

CATTGTATAT CATTGCGCTG GCATTTGGAG TGATTGCAAC AATTATAACT GCCACAATCA
1260

GTTATGTATT TTCAACACAA ATTACTAAAC CGCTTGTCAG TTTATCAAAT AAAATGATTG
1320

AGATTCGACG AGATGGTTTT CAAAATAAAT TGCAATTAAA TACAAATTAT GAAGAAATAG
1380

ATAATTTAGC AAATACGTTT AATGAGATGA TGAGCCAAAT TGAAGAATCA TTTAATCAAC
1440

AAAGACAATT TGTTGAAGAT GCGTCACATG AATTACGAAC ACCATTACAA ATTATTCAAG
1500

GTCATTTAAA TTTGATTCAG CGATGGGGAA AAAAAGACCC AGCAGTATTA GAAGAATCGT
1560

TAAATATTTC TATTGAAGAA ATGAATCGTA TCATAAAATT AGTCGAAGAA TTACTTGAAT
1620
TGACTAAAGG AGATGTAAAT GACATTTCTT CTGAAGCGCA GACCGTGCAT ATTAATGATG
1680
AAATTCGCTC GCGAATACAC TCATTAAAAC AATTGCATCC TGATTATCAA TTTGATACGG
1740
ATCTGACATC TAAAAATCTA GAAATTAAAA TGAAACCTCA TCAATTCGAA CAATTATTTT
1800
TAATCTTTAT TGATAATGCA ATCAAATATG ATGTGAAGAA TAAGAAAATT AAAGTTAAGA
1860
CAAGGTTAAA AAATAAGCAA AAAATAATTG AAATTACAGA TCATGGAATT GGTATTCCAG
1920
AGGAAGATCA AGATTTCATT TTTGATCGCT TTTATCGAGT GGATAAATCT CGTTCAAGAA
1980
GTCAAGGCGG TAATGGACTC GGATTATCTA TTGCTCAAAA AATCATTCAA TTAAACGGAG
2040
GATCGATTAA AATTAAAAGT GAAATTAATA AAGGAACAAC GTTTAAAATC ATATTTTAAT
2100
CATGTCTGAG ACGTCAATCA AAGTCATAGG ATCAATTTTT TAAGTACACA TTAGCTGTGA
2160
CTAATGTATA AGAACAACTA TAAAACAAAT AAACAGTGGT T
2201

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 451 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Thr Lys Arg Lys Leu Arg Asn Asn Trp Ile Ile Val Thr Thr Met
1 5 10 15
Ile Thr Phe Val Thr Ile Phe Leu Phe Cys Leu Ile Ile Ile Phe Phe
20 25 30
Leu Lys Asp Thr Leu His Asn Ser Glu Leu Asp Asp Ala Glu Arg Ser
35 40 45

Ser Ser Asp Ile Asn Asn Leu Phe His Ser Lys Pro Val Lys Asp Ile
50 55 60

Ser Ala Leu Asp Leu Asn Ala Ser Leu Gly Asn Phe Gln Glu Ile Ile
65 70 75 80

Ile Tyr Asp Glu His Asn Asn Lys Leu Phe Glu Thr Ser Asn Asp Asn
85 90 95

Thr Val Arg Val Glu Pro Gly Tyr Glu His Arg Tyr Phe Asp Arg Val
100 105 110

Ile Lys Lys Arg Tyr Lys Gly Ile Glu Tyr Leu Ile Ile Lys Glu Pro
115 120 125

Ile Thr Thr Gln Asp Phe Lys Gly Tyr Ser Leu Leu Ile His Ser Leu
130 135 140

Glu Asn Tyr Asp Asn Ile Val Lys Ser Leu Tyr Ile Ile Ala Leu Ala
145 150 155 160

Phe Gly Val Ile Ala Thr Ile Ile Thr Ala Thr Ile Ser Tyr Val Phe
165 170 175

Ser Thr Gln Ile Thr Lys Pro Leu Val Ser Leu Ser Asn Lys Met Ile
180 185 190

Glu Ile Arg Arg Asp Gly Phe Gln Asn Lys Leu Gln Leu Asn Thr Asn
195 200 205

Tyr Glu Glu Ile Asp Asn Leu Ala Asn Thr Phe Asn Glu Met Met Ser
210 215 220

Gln Ile Glu Glu Ser Phe Asn Gln Gln Arg Gln Phe Val Glu Asp Ala
225 230 235 240

Ser His Glu Leu Arg Thr Pro Leu Gln Ile Ile Gln Gly His Leu Asn
245 250 255

Leu Ile Gln Arg Trp Gly Lys Lys Asp Pro Ala Val Leu Glu Glu Ser
260 265 270

Leu Asn Ile Ser Ile Glu Glu Met Asn Arg Ile Ile Lys Leu Val Glu
275 280 285

Glu Leu Leu Glu Leu Thr Lys Gly Asp Val Asn Asp Ile Ser Ser Glu
290 295 300

Ala Gln Thr Val His Ile Asn Asp Glu Ile Arg Ser Arg Ile His Ser
305 310 315 320

Leu Lys Gln Leu His Pro Asp Tyr Gln Phe Asp Thr Asp Leu Thr Ser
325 330 335

Lys Asn Leu Glu Ile Lys Met Lys Pro His Gln Phe Glu Gln Leu Phe
340 345 350

Leu Ile Phe Ile Asp Asn Ala Ile Lys Tyr Asp Val Lys Asn Lys Lys
355 360 365

Ile Lys Val Lys Thr Arg Leu Lys Asn Lys Gln Lys Ile Ile Glu Ile
370 375 380

Thr Asp His Gly Ile Gly Ile Pro Glu Glu Asp Gln Asp Phe Ile Phe
385 390 395 400

Asp Arg Phe Tyr Arg Val Asp Lys Ser Arg Ser Arg Ser Gln Gly Gly
405 410 415

Asn Gly Leu Gly Leu Ser Ile Ala Gln Lys Ile Ile Gln Leu Asn Gly
420 425 430

Gly Ser Ile Lys Ile Lys Ser Glu Ile Asn Lys Gly Thr Thr Phe Lys
435 440 445

Ile Ile Phe
450

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 736 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

ATTTACGTTT TGTCATCGTA TCACATACCC AACGCCACGA ACTGTTTCAA TCATTTTGTC
60

ACGATCGTAT GGTTTTAACT TGTTTCGTAA ATATCTTATA TAAACATCTA CGACATTTGT
120

TTCTACTTCA CTATTATAAC CCCATACATG ATTTAAAATT TGTTCCCGTT GCATAACATG
180

GTTTTTATTT TCAGCTAGAA GATATAGTAA ATCATACTCT GTTTTTGTTA ATTCAATTTC
240

TGCGCCATTT ACCGTCACTT TAAAAGCGTT CTTATCAATT GTAATACCGT TGACATCGAT
300

AATATCCTTT TGTGGCTGAC GACGTAAAAT TGCACGAATT CTTGCTAAAA GTTCTTCAAT
360

ATCAAACGGC TTAACTATAT AATCGTCTGC ACCGTAATCA AGCCCAGCAA CTTTGTCATA
420

CGTATCACTT TTCGCTGTAA TTATAATGAT AGGTGTAGAT TGTTGTTGTC TAATTTTGCG
480

ACAAATTTCT AAGCCATTAA TTGACGGCAA CATTAAATCT AATATGATTA AATCATAGTA
540
ATGGCTAAGC GCTTTATCTA AACCGTCTTG TCCATCATAC TCTGTGTCCA CATTGTAATT
600
TTCATGTGTG AGTTCCAATT CAAGAAATCT TGCTAAGTTT TGTTCATCTT CTACTATTAA
660
AATTTGCGTC ATTTGTACAC CTCATATTAC GACTTTTTCT AATAAGGTAA TATATTTTAA
720
ATTTTGTATT TTTCTA
736

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 219 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Thr Gln Ile Leu Ile Val Glu Asp Glu Gln Asn Leu Ala Arg Phe
1 5 10 15
Leu Glu Leu Glu Leu Thr His Glu Asn Tyr Asn Val Asp Thr Glu Tyr
20 25 30
Asp Gly Gln Asp Gly Leu Asp Lys Ala Leu Ser His Tyr Tyr Asp Leu
35 40 45
Ile Ile Leu Asp Leu Met Leu Pro Ser Ile Asn Gly Leu Glu Ile Cys
50 55 60
Arg Lys Ile Arg Gln Gln Gln Ser Thr Pro Ile Ile Ile Ile Thr Ala
65 70 75 80
Lys Ser Asp Thr Tyr Asp Lys Val Ala Gly Leu Asp Tyr Gly Ala Asp
85 90 95
Asp Tyr Ile Val Lys Pro Phe Asp Ile Glu Glu Leu Leu Ala Arg Ile
100 105 110
Arg Ala Ile Leu Arg Arg Gln Pro Gln Lys Asp Ile Ile Asp Val Asn
115 120 125
Gly Ile Thr Ile Asp Lys Asn Ala Phe Lys Val Thr Val Asn Gly Ala
130 135 140

```
Glu Ile Glu Leu Thr Lys Thr Glu Tyr Asp Leu Leu Tyr Leu Leu Ala
145                 150                 155                 160
Glu Asn Lys Asn His Val Met Gln Arg Glu Gln Ile Leu Asn His Val
                165                 170                 175
Trp Gly Tyr Asn Ser Glu Val Glu Thr Asn Val Val Asp Val Tyr Ile
            180                 185                 190
Arg Tyr Leu Arg Asn Lys Leu Lys Pro Tyr Asp Arg Asp Lys Met Ile
        195                 200                 205
Glu Thr Val Arg Gly Val Gly Tyr Val Ile Arg
    210                 215
```

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
ATGACAAAAC GTAAATTGCG CAATAAC
27
```

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
AAATATGATT TTAAACGTTG TTCC
24
```

What is claimed is:

1. An isolated polynucleotide comprising a polynucleotide having at least a 70% identity to a polynucleotide encoding a polypeptide comprising the amino acid sequence of SEQ ID NO:2.

2. An isolated polynucleotide comprising a polynucleotide having at least a 70% identity to a polynucleotide encoding the same mature polypeptide expressed by the Histidine Kinase gene contained in the *Staphylococcus aureus*.

3. An isolated polynucleotide comprising a polynucleotide encoding a polypeptide comprising an amino acid sequence which is at least 70% identical to the amino acid sequence of SEQ ID NO:2.

4. An isolated polynucleotide that is complementary to the polynucleotide of claim 1

5. The polynucleotide of Claim 1 wherein the polynucleotide is DNA or RNA

6. The polynucleotide of Claim 1 comprising the nucleic acid sequence set forth in SEQ ID NO:1.

7. The polynucleotide of Claim 1 comprising nucleotide 744 to the stop codon which begins at nucleotide number 2097 set forth in SEQ ID NO:1.

8. The polynucleotide of Claim 1 which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:2.

9. A vector comprising the polynucleotide of Claim 1.

10. A host cell comprising the vector of Claim 9.

11. A process for producing a polypeptide comprising: expressing from the host cell of Claim 10 a polypeptide encoded by said DNA.

12. A process for producing a Histidine Kinase polypeptide or fragment comprising culturing a host of claim 10 under conditions sufficient for the production of said polypeptide or fragment.

13. A polypeptide comprising an amino acid sequence which is at least 70% identical to the amino acid sequence of SEQ ID NO:2.

14. A polypeptide comprising an amino acid sequence as set forth in SEQ ID NO:2.

15. An antibody against the polypeptide of claim 14.

16. An antagonist which inhibits the activity or expression of the polypeptide of claim 14.

17. A method for the treatment of an individual in need of Histidine Kinase polypeptide comprising: administering to the individual a therapeutically effective amount of the polypeptide of claim 14.

18. A method for the treatment of an individual having need to inhibit Histidine Kinase polypeptide comprising: administering to the individual a therapeutically effective amount of the antagonist of Claim 14.

19. A process for diagnosing a disease related to expression or activity of the polypeptide of claim 14 in an individual comprising:

(a) determining a nucleic acid sequence encoding said polypeptide, and/or

(b) analyzing for the presence or amount of said polypeptide in a sample derived from the individual.

20. A method for identifying compounds which interact with and inhibit or activate an activity of the polypeptide of claim 14 comprising:

contacting a composition comprising the polypeptide with the compound to be screened under conditions to permit interaction between the compound and the polypeptide to assess the interaction of a compound, such interaction being associated with a second component capable of providing a detectable signal in response to the interaction of the polypeptide with the compound;

and determining whether the compound interacts with and activates or inhibits an activity of the polypeptide by detecting the presence or absence of a signal generated from the interaction of the compound with the polypeptide.

21. A method for inducing an immunological response in a mammal which comprises inoculating the mammal with Histidine Kinase polypeptide of claim 14, or a fragment or variant thereof, adequate to produce antibody and/or T cell immune response to protect said animal from disease.

22. A method of inducing immunological response in a mammal which comprises delivering a nucleic acid vector to direct expression of Histidine Kinase polypeptide of claim 14, or fragment or a variant thereof, for expressing said Histidine Kinase polypeptide, or a fragment or a variant thereof *in vivo* in order to induce an immunological response to produce antibody and/ or T cell immune response to protect said animal from disease.

23. An isolated polynucleotide comprising a polynucleotide having at least a 70% identity to a polynucleotide encoding a polypeptide comprising the amino acid sequence of SEQ ID NO:4.

24. An isolated polynucleotide comprising a polynucleotide having at least a 70% identity to the polynucleotide sequence of SEQ ID NO:3.

ABSTRACT OF THE DISCLOSURE

The invention provides Histidine Kinase polypeptides and polynucleotides encoding Histidine Kinase polypeptides and methods for producing such polypeptides by recombinant techniques. Also provided are methods for utilizing Histidine Kinase polypeptides to screen for antibacterial compounds.

This Raw Listing contains the General Information Section and up to the first 5 pages.

SEQUENCE LISTING

(1) General Information

(i) APPLICANT: Wallis, Nicola G.
Shilling, Lisa K.
Mooney, Jeffrey L.
Debouck, Christine
Zhong, YiYi
Jaworski, Deborah D.
Wang, Min
Throup, John P.

(ii) TITLE OF THE INVENTION: Histidine Kinase

(iii) NUMBER OF SEQUENCES: 6

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Dechert, Price & Rhoads
(B) STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
(C) CITY: Philadelphia
(D) STATE: PA
(E) COUNTRY: USA
(F) ZIP: 19103-2793

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Diskette
(B) COMPUTER: IBM Compatible
(C) OPERATING SYSTEM: DOS
(D) SOFTWARE: FastSEQ for Windows Version 2.0

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Falk, Stephen T
(B) REGISTRATION NUMBER: 36,795

(C) REFERENCE/DOCKET NUMBER: GM10127

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 215-994-2488
(B) TELEFAX: 215-994-2222
(C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2201 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
TAATTTAAAA AGCAACTATT GTATAGAAAA ATACAAAATT TAAAATATAT TACCTTATTA     60
GAAAAAGTCG TAATATGAGG TGTACAAATG ACGCAAATTT TAATAGTAGA AGATGAACAA    120
AACTTAGCAA GATTTCTTGA ATTGGAACTC ACACATGAAA ATTACAATGT GGACACAGAG    180
TATGATGGAC AAGACGGTTT AGATAAAGCG CTTAGCCATT ACTATGATTT AATCATATTA    240
GATTTAATGT TGCCGTCAAT TAATGGCTTA GAAATTTGTC GCAAAATTAG ACAACAACAA    300
TCTACACCTA TCATTATAAT TACAGCGAAA AGTGATACGT ATGACAAAGT TGCTGGGCTT    360
GATTACGGTG CAGACGATTA TATAGTTAAG CCGTTTGATA TTGAAGAACT TTTAGCAAGA    420
ATTCGTGCAA TTTTACGTCG TCAGCCACAA AAGGATATTA TCGATGTCAA CGGTATTACA    480
ATTGATAAGA ACGCTTTTAA AGTGACGGTA AATGGCGCAG AAATTGAATT AACAAAAACA    540
GAGTATGATT TACTATATCT TCTAGCTGAA AATAAAAACC ATGTTATGCA ACGGGAACAA    600
ATTTTAAATC ATGTATGGGG TTATAATAGT GAAGTAGAAA CAAATGTCGT AGATGTTTAT    660
ATAAGATATT TACGAAACAA GTTAAAACCA TACGATCGTG ACAAAATGAT TGAAACAGTT    720
CGTGGCGTTG GGTATGTGAT ACGATGACAA AACGTAAATT GCGCAATAAC TGGATTATTG    780
TTACCACGAT GATTACGTTT GTCACGATAT TTTTGTTTTG TTTAATTATT ATTTTTTTCT    840
TGAAAGATAC ACTGCATAAT AGTGAGCTTG ATGATGCAGA ACGAAGCTCA AGCGATATTA    900
ATAATTTATT TCATTCTAAG CCTGTTAAAG ATATATCTGC ATTAGACTTG AATGCATCTT    960
TAGGTAATTT TCAAGAGATA ATTATTTATG ATGAGCATAA TAATAAATTA TTTGAGACAT   1020
CGAATGATAA CACAGTGAGA GTTGAACCAG GTTATGAACA CCGTTATTTT GACCGCGTAA   1080
TAAAAAAACG CTATAAAGGC ATTGAATATT TAATTATTAA AGAACCAATT ACAACGCAAG   1140
ATTTCAAAGG GTATAGCTTG TTAATTCATT CACTAGAAAA TTATGATAAC ATCGTAAAAT   1200
CATTGTATAT CATTGCGCTG GCATTTGGAG TGATTGCAAC AATTATAACT GCCACAATCA   1260
GTTATGTATT TTCAACACAA ATTACTAAAC CGCTTGTCAG TTTATCAAAT AAAATGATTG   1320
AGATTCGACG AGATGGTTTT CAAAATAAAT TGCAATTAAA TACAAATTAT GAAGAAATAG   1380
ATAATTTAGC AAATACGTTT AATGAGATGA TGAGCCAAAT TGAAGAATCA TTTAATCAAC   1440
AAAGACAATT TGTTGAAGAT GCGTCACATG AATTACGAAC ACCATTACAA ATTATTCAAG   1500
GTCATTTAAA TTTGATTCAG CGATGGGGAA AAAAAGACCC AGCAGTATTA GAAGAATCGT   1560
TAAATATTTC TATTGAAGAA ATGAATCGTA TCATAAAATT AGTCGAAGAA TTACTTGAAT   1620
TGACTAAAGG AGATGTAAAT GACATTTCTT CTGAAGCGCA GACCGTGCAT ATTAATGATG   1680
AAATTCGCTC GCGAATACAC TCATTAAAAC AATTGCATCC TGATTATCAA TTTGATACGG   1740
ATCTGACATC TAAAAATCTA GAAATTAAAA TGAAACCTCA TCAATTCGAA CAATTATTTT   1800
TAATCTTTAT TGATAATGCA ATCAAATATG ATGTGAAGAA TAAGAAAATT AAAGTTAAGA   1860
CAAGGTTAAA AAATAAGCAA AAAATAATTG AAATTACAGA TCATGGAATT GGTATTCCAG   1920
AGGAAGATCA AGATTTCATT TTTGATCGCT TTTATCGAGT GGATAAATCT CGTTCAAGAA   1980
GTCAAGGCGG TAATGGACTC GGATTATCTA TTGCTCAAAA AATCATTCAA TTAAACGGAG   2040
```

```
GATCGATTAA AATTAAAAGT GAAATTAATA AAGGAACAAC GTTTAAAATC ATATTTTAAT   2100
CATGTCTGAG ACGTCAATCA AAGTCATAGG ATCAATTTTT TAAGTACACA TTAGCTGTGA   2160
CTAATGTATA AGAACAACTA TAAAACAAAT AAACAGTGGT T                       2201
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 451 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Thr Lys Arg Lys Leu Arg Asn Asn Trp Ile Ile Val Thr Thr Met
 1               5                  10                  15
Ile Thr Phe Val Thr Ile Phe Leu Phe Cys Leu Ile Ile Ile Phe Phe
            20                  25                  30
Leu Lys Asp Thr Leu His Asn Ser Glu Leu Asp Asp Ala Glu Arg Ser
        35                  40                  45
Ser Ser Asp Ile Asn Asn Leu Phe His Ser Lys Pro Val Lys Asp Ile
    50                  55                  60
Ser Ala Leu Asp Leu Asn Ala Ser Leu Gly Asn Phe Gln Glu Ile Ile
65                  70                  75                  80
Ile Tyr Asp Glu His Asn Asn Lys Leu Phe Glu Thr Ser Asn Asp Asn
                85                  90                  95
Thr Val Arg Val Glu Pro Gly Tyr Glu His Arg Tyr Phe Asp Arg Val
            100                 105                 110
Ile Lys Lys Arg Tyr Lys Gly Ile Glu Tyr Leu Ile Ile Lys Glu Pro
        115                 120                 125
Ile Thr Thr Gln Asp Phe Lys Gly Tyr Ser Leu Leu Ile His Ser Leu
    130                 135                 140
Glu Asn Tyr Asp Asn Ile Val Lys Ser Leu Tyr Ile Ile Ala Leu Ala
145                 150                 155                 160
Phe Gly Val Ile Ala Thr Ile Ile Thr Ala Thr Ile Ser Tyr Val Phe
                165                 170                 175
Ser Thr Gln Ile Thr Lys Pro Leu Val Ser Leu Ser Asn Lys Met Ile
            180                 185                 190
Glu Ile Arg Arg Asp Gly Phe Gln Asn Lys Leu Gln Leu Asn Thr Asn
        195                 200                 205
Tyr Glu Glu Ile Asp Asn Leu Ala Asn Thr Phe Asn Glu Met Met Ser
    210                 215                 220
Gln Ile Glu Glu Ser Phe Asn Gln Gln Arg Gln Phe Val Glu Asp Ala
225                 230                 235                 240
Ser His Glu Leu Arg Thr Pro Leu Gln Ile Ile Gln Gly His Leu Asn
                245                 250                 255
Leu Ile Gln Arg Trp Gly Lys Lys Asp Pro Ala Val Leu Glu Glu Ser
            260                 265                 270
Leu Asn Ile Ser Ile Glu Glu Met Asn Arg Ile Ile Lys Leu Val Glu
        275                 280                 285
Glu Leu Leu Glu Leu Thr Lys Gly Asp Val Asn Asp Ile Ser Ser Glu
    290                 295                 300
```

```
Ala Gln Thr Val His Ile Asn Asp Glu Ile Arg Ser Arg Ile His Ser
305                 310                 315                 320
Leu Lys Gln Leu His Pro Asp Tyr Gln Phe Asp Thr Asp Leu Thr Ser
                325                 330                 335
Lys Asn Leu Glu Ile Lys Met Lys Pro His Gln Phe Glu Gln Leu Phe
            340                 345                 350
Leu Ile Phe Ile Asp Asn Ala Ile Lys Tyr Asp Val Lys Asn Lys Lys
        355                 360                 365
Ile Lys Val Lys Thr Arg Leu Lys Asn Lys Gln Lys Ile Ile Glu Ile
    370                 375                 380
Thr Asp His Gly Ile Gly Ile Pro Glu Glu Asp Gln Asp Phe Ile Phe
385                 390                 395                 400
Asp Arg Phe Tyr Arg Val Asp Lys Ser Arg Ser Arg Ser Gln Gly Gly
                405                 410                 415
Asn Gly Leu Gly Leu Ser Ile Ala Gln Lys Ile Ile Gln Leu Asn Gly
            420                 425                 430
Gly Ser Ile Lys Ile Lys Ser Glu Ile Asn Lys Gly Thr Thr Phe Lys
        435                 440                 445
Ile Ile Phe
    450
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 736 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
ATTTACGTTT TGTCATCGTA TCACATACCC AACGCCACGA ACTGTTTCAA TCATTTTGTC     60
ACGATCGTAT GGTTTTAACT TGTTTCGTAA ATATCTTATA TAAACATCTA CGACATTTGT    120
TTCTACTTCA CTATTATAAC CCCATACATG ATTTAAAATT TGTTCCCGTT GCATAACATG    180
GTTTTTATTT TCAGCTAGAA GATATAGTAA ATCATACTCT GTTTTTGTTA ATTCAATTTC    240
TGCGCCATTT ACCGTCACTT TAAAAGCGTT CTTATCAATT GTAATACCGT TGACATCGAT    300
AATATCCTTT TGTGGCTGAC GACGTAAAAT TGCACGAATT CTTGCTAAAA GTTCTTCAAT    360
ATCAAACGGC TTAACTATAT AATCGTCTGC ACCGTAATCA AGCCCAGCAA CTTTGTCATA    420
CGTATCACTT TTCGCTGTAA TTATAATGAT AGGTGTAGAT TGTTGTTGTC TAATTTTGCG    480
ACAAATTTCT AAGCCATTAA TTGACGGCAA CATTAAATCT AATATGATTA AATCATAGTA    540
ATGGCTAAGC GCTTTATCTA AACCGTCTTG TCCATCATAC TCTGTGTCCA CATTGTAATT    600
TTCATGTGTG AGTTCCAATT CAAGAAATCT TGCTAAGTTT TGTTCATCTT CTACTATTAA    660
AATTTGCGTC ATTTGTACAC CTCATATTAC GACTTTTTCT AATAAGGTAA TATATTTTAA    720
ATTTTGTATT TTTCTA                                                    736
```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 219 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

```
206
207
208         (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:
209
210    Met Thr Gln Ile Leu Ile Val Glu Asp Glu Gln Asn Leu Ala Arg Phe
211     1               5                  10                  15
212    Leu Glu Leu Glu Leu Thr His Glu Asn Tyr Asn Val Asp Thr Glu Tyr
213                20                  25                  30
214    Asp Gly Gln Asp Gly Leu Asp Lys Ala Leu Ser His Tyr Tyr Asp Leu
215            35                  40                  45
216    Ile Ile Leu Asp Leu Met Leu Pro Ser Ile Asn Gly Leu Glu Ile Cys
217        50                  55                  60
218    Arg Lys Ile Arg Gln Gln Gln Ser Thr Pro Ile Ile Ile Ile Thr Ala
219    65                  70                  75                  80
220    Lys Ser Asp Thr Tyr Asp Lys Val Ala Gly Leu Asp Tyr Gly Ala Asp
221                    85                  90                  95
222    Asp Tyr Ile Val Lys Pro Phe Asp Ile Glu Glu Leu Leu Ala Arg Ile
223                100                 105                 110
224    Arg Ala Ile Leu Arg Arg Gln Pro Gln Lys Asp Ile Ile Asp Val Asn
225            115                 120                 125
226    Gly Ile Thr Ile Asp Lys Asn Ala Phe Lys Val Thr Val Asn Gly Ala
227        130                 135                 140
228    Glu Ile Glu Leu Thr Lys Thr Glu Tyr Asp Leu Leu Tyr Leu Leu Ala
229    145                 150                 155                 160
230    Glu Asn Lys Asn His Val Met Gln Arg Glu Gln Ile Leu Asn His Val
231                    165                 170                 175
232    Trp Gly Tyr Asn Ser Glu Val Glu Thr Asn Val Val Asp Val Tyr Ile
233                180                 185                 190
234    Arg Tyr Leu Arg Asn Lys Leu Lys Pro Tyr Asp Arg Asp Lys Met Ile
235            195                 200                 205
236    Glu Thr Val Arg Gly Val Gly Tyr Val Ile Arg
237        210                 215
238
239             (2) INFORMATION FOR SEQ ID NO:5:
240
241           (i) SEQUENCE CHARACTERISTICS:
242             (A) LENGTH: 27 base pairs
243             (B) TYPE: nucleic acid
244             (C) STRANDEDNESS: single
245             (D) TOPOLOGY: linear
246
247
248           (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:
249
250    ATGACAAAAC GTAAATTGCG CAATAAC                                      27
251
252             (2) INFORMATION FOR SEQ ID NO:6:
253
254           (i) SEQUENCE CHARACTERISTICS:
255             (A) LENGTH: 24 base pairs
256             (B) TYPE: nucleic acid
257             (C) STRANDEDNESS: single
258             (D) TOPOLOGY: linear
```

SEQUENCE VERIFICATION REPORT
PATENT APPLICATION *US/09/006,627*

DATE: 02/19/98
TIME: 18:26:45

INPUT SET: S23574.raw

Line	Error	Original Text